

Resum

Aquest projecte s'emmarca dins del que es coneix com a "iniciatives Open Data". L'Open Data (anglès per dades obertes) consisteix en la publicació de dades de forma oberta. La publicació de conjunts de dades al públic per part d'administracions i cada cop més entitats no només fomenta la transparència a nivell de processos i activitats d'una institució, sinó que dóna l'oportunitat de crear projectes o aplicacions que molts cops es converteixen en negocis, i dels quals se'n poden beneficiar tant administracions com ciutadans.

En aquest projecte, doncs, s'ha contextualitzat la situació actual de l'Open Data i s'ha dut a terme una recerca sobre els portals capdavaners en aquest apartat, tant a àmbit local i autonòmic, com a àmbit nacional i internacional. Aquesta recerca està enfocada a explorar les característiques d'aquests portals, quin tipus de dades presenten, i fins a quin punt està estès l'Open Data.

En segon lloc, s'han cercat portals que no es limitin a aportar catàlegs de dades, sinó que en facin una explotació visual a mode d'aplicació (seguint la filosofia que posseir moltes dades no és sinònim de més coneixement, si no es saben tractar correctament) .

Aquesta cerca ha estat la precursora del propòsit final d'aquest projecte: el de realitzar una aplicació o interfície visual de creació pròpia. Per fer-ho s'ha fet un estudi sobre les opcions més idònies per elaborar una aplicació d'aquest tipus, del qual n'ha sortit vencedor la plataforma Shiny, vinculada al llenguatge de programació R.

La interfície visual s'ha creat en forma d'explorador socioeconòmic, pel qual s'han necessitat bases de dades que s'han obtingut d'un dels portals recercats prèviament (el del Banc Mundial).

Finalment, s'ha creat amb èxit l'aplicació, donant una visió més acurada de quin tipus de projecte es pot dur a terme a partir de l'explotació de dades obertes, i proporcionant la certesa també, que el moviment Open Data no està fent més que créixer, i les possibilitats que presenta són moltes i molt interessants.



Sumari

RESUM	1
SUMARI	3
1. PREFACI	5
1.1. Origen del projecte	5
1.2. Motivació	5
2. INTRODUCCIÓ	7
2.1. Objectius i abast del projecte	7
3. INTRODUCCIÓ A L'OPEN DATA	9
3.1. Concepte Open Data.....	9
3.2. Decàleg d'obertura	9
3.3. Objectius.....	10
3.4. Beneficis	10
4. ESTAT DE L'ART SOBRE LA DISPONIBILITAT I ÚS DE DADES	13
4.1. Marc tecnològic	13
4.1.1. Estructura i temàtiques	13
4.1.2. Formats.....	14
4.1.3. Tim Berners-Lee i les seves 5 estrelles.....	16
4.2. Llistat de recursos en l'àmbit espanyol.....	18
4.2.1. Nacional.....	19
4.2.2. Autonòmic.....	20
4.2.3. Local	21
4.3. Llistat de recursos en l'àmbit internacional.....	23
4.4. Portals que exploten les dades de manera visual i interactiva.....	26
5. ESTUDI DE LES OPCIONS PER FER UNA INTERFÍCIE D'EXPLOTACIÓ DE DADES	32
5.1. Python	32
5.2. Flash.....	33
5.3. Shiny de RStudio.....	34
5.4. Decisió.....	36
6. EXPLOTACIÓ I REPRESENTACIÓ D'UNA SÈRIE DE DADES	39

6.1. Obtenció de les dades	39
6.2. Preparació de les dades	43
7. CREACIÓ DE LA INTERFÍCIE	47
7.1. Introducció a Shiny	47
7.2. Desenvolupament i creació de la interfície	51
7.2.1. Explorador de dades.....	51
7.2.2. Anàlisi de relacions	53
7.2.3. Anàlisi geogràfic.....	56
8. EXPLOTACIÓ DE LA INTERFÍCIE	60
9. CONCLUSIONS	64
9.1. Revisió dels objectius plantejats	64
9.2. Línies d'evolució futures de l'aplicació	65
9.3. Revisió crítica de la feina feta	66
AGRAÏMENTS	67
BIBLIOGRAFIA.....	68
Referències bibliogràfiques	68
ANNEX	70



1. Prefaci

1.1. Origen del projecte

L'origen del projecte sorgeix a partir de la voluntat d'estudiar i aprofundir sobre un tema que fos innovador pel que fa a l'estadística d'avui en dia. Davant d'això s'ha tingut en compte el creixement recent de l'anomenat "moviment Open Data" pel qual les administracions i empreses estant posant a l'ús del públic les dades de les quals disposen. El seu propòsit no és tan sols el de millorar la transparència dels processos que duen a terme sinó també la creació de valor afegit, perquè estan adonant-se que la reutilització i l'explotació d'aquestes dades pot ser una font de beneficis molt importants, tant a nivell econòmic com a nivell d'avenços tecnològics.

Davant d'això, centenars d'institucions i entitats arreu del món estan creant portals on publiquen dades obertes i duen a terme iniciatives de rellevància a partir de l'ús i l'explotació de dades obertes.

Aquest projecte, doncs, va néixer amb la idea d'explorar aquestes iniciatives, intentar descobrir quin abast tenen els projectes que es poden dur a terme a partir de la explotació i la reutilització de dades, i esbrinar si seria possible elaborar un projecte de creació pròpia a partir de bases de dades obertes.

1.2. Motivació

La motivació per fer aquest projecte va sorgir per diversos motius.

El primer és ben senzill. Des de sempre m'ha agradat tot allò relacionat amb l'estadística. Les dues assignatures d'estadística que he cursat a l'ETSEIB són dues de les que més m'han agradat (Estadística i Tècniques Estadístiques per a la Qualitat). Aquestes dues assignatures em van permetre entendre molts aspectes de l'estadística que desconeixia i profunditzar sobre d'altres. En particular, em va interessar molt i vaig gaudir fent el treball de curs que s'havia de realitzar en l'assignatura de Tècniques Estadístiques per a la Qualitat.

De seguida vaig pensar que volia que el meu TFG hi tingués a veure, així que vaig contactar amb el professor del departament d'estadística de l'escola, i coordinador de Tècniques Estadístiques per a la Qualitat, el professor Lluís Marco Almagro. Ell va ser qui em va introduir en el concepte de l'Open Data i entre els dos vam decidir el tema del qual tractaria aquest projecte.

A més a més, fa temps que sento curiositat per la programació i disseny d'aplicacions webs però mai he tingut realment temps per dedicar-m'hi. El fet de poder crear-ne una que a més tingués a veure amb conceptes estadístics em va entusiasmar des del principi i em vaig fixar l'objectiu que en un futur els coneixements adquirits en aquest projecte servissin de base per seguir aprenent i poder crear aplicacions cada cop de dificultat més elevada.

2. Introducció

2.1. Objectius i abast del projecte

Aquest projecte consisteix en analitzar la situació actual de l'Open Data, donat que aquest moviment cada cop té més rellevància per a la societat, i el nombre d'iniciatives que se'n deriven creix dia rere dia. Es pretén introduir-se en el moviment, per entendre quina importància té, quins avantatges planteja i quins beneficis aporta a la ciutadania i a les administracions.

Un cop contextualitzada la situació actual de l'Open Data es vol fer una recerca sobre els portals que estan creant iniciatives a partir de dades obertes tant a àmbit nacional com internacional, i estudiar-ne les característiques. Amb aquesta recerca s'intentarà calibrar la magnitud del moviment Open Data i descobrir si està restringit a països capdavaners o si en canvi s'està estenent fins i tot a àmbit local.

El sol fet de recopilar dades no és sinònim de tenir més informació, sinó que la majoria de cops és necessari anar més enllà i representar aquestes dades de la manera més idònia per extreure'n les conclusions necessàries. Per tant, el següent objectiu d'aquest projecte és el d'indagar sobre quines opcions existeixen a nivell de representació de dades i explorar si hi ha portals que a part de tenir catàlegs de dades obertes es dediquin a dur a terme interfícies que permetin la interpretació d'aquestes dades de forma visual i sense requerir que els usuaris tinguin coneixements d'estadística.

D'aquesta manera, aprofitant la recerca feta, es recopilaran conjunts de dades de rellevància socioeconòmica d'un dels portals capdavaners en matèria d'Open Data per tal de complir el següent objectiu. Aquest és potser el de més envergadura, ja que no es limita a la recerca sinó que consisteix en provar de crear una interfície visual pròpia, a mode d'aplicació web, per a l'explotació d'aquestes dades mitjançant eines estadístiques.

Per aconseguir-ho caldrà investigar i fer l'estudi corresponent sobre quina és l'eina òptima per a l'elaboració d'una aplicació d'aquest caire (sens dubte s'haurà de recórrer a la programació). Aquesta interfície ha de permetre a l'usuari interactuar amb ella de diverses maneres, de forma que pugui arribar a conclusions clares visualment pel que fa als conjunts de dades recopilades sense necessitat de tenir coneixements previs en estadística.

L'elaboració d'una interfície com la que es vol crear presenta moltes possibilitats pel que fa a les dades a representar. A més a més, les entitats que posen a l'abast del públic els seus catàlegs de dades disposen de centenars de conjunts de dades agrupades en temàtiques molt diverses. Tanmateix, aquest projecte es centra en dades de rellevància

socioeconòmica, ja que s'ha considerat que aquest tipus de dades permetran una representació visual fluïda i serà possible fer interpretacions que tinguin un interès general.

3. Introducció a l'Open Data

3.1. Concepte Open Data

L'*Open Data Institute* defineix Open Data (o dades obertes, en català), com aquelles dades que organitzacions, empreses, o individuals han posat a l'abast del públic perquè tothom hi pugui accedir, les pugui utilitzar o fins i tot compartir. (Berners-Lee, 2012)

Aquestes dades han d'estar disponibles de manera lliure sense cap tipus de restricció, ja siguin per patents, drets d'autor o altres mecanismes de control.

3.2. Decàleg d'obertura

És important poder assegurar que s'està parlant de dades obertes i no d'algun altre concepte similar però que no entraria dins de la filosofia Open Data talment dita. Per assegurar-ho les dades ofertes han de complir els principis exposats en el següent decàleg:

- **Públiques:** totes aquelles dades de caràcter públic han d'obrir-se, a no ser que estiguin sotmeses a restriccions de drets d'autor, privacitat o seguretat. D'aquesta manera s'evitaria qualsevol debat per part de l'administració a l'hora de decidir quines dades s'han de publicar i quines no.
- **Estructurades:** Les dades han d'estar publicades "en brut", és a dir, tal i com estaven en el seu origen, sense processar i mantenint el màxim nivell de detall possible.
- **Actualitzades:** les dades han de ser actuals i precises tal que no perdin el seu valor, i per tant caldria posar-les a disposició del públic en la freqüència necessària.
- **Accessibles:** les dades haurien de poder arribar a tothom qui les necessiti, sense cap mena de restricció per al seu ús.
- **Automatitzades:** les dades han d'estar estructurades de manera que un ordinador pugui processar-les automàticament.
- **Sense registre:** han d'estar disponibles sense necessitat que l'usuari que en vulgui disposar s'hagi d'identificar.
- **Obertes:** en el sentit de no propietat. No poden dependre d'una entitat o una eina propietària d'una entitat. Per exemple, quan parlem de format obert parlem de CSV o

XML, i formats propietaris serien bé Word o Excel.

- **Lliures:** les dades que siguin considerades obertes han d'estar lliures de copyright o patents i no estar sotmesos a drets de privacitat, seguretat o privilegis.

3.3. Objectius

L'obertura de dades persegueix molts objectius, entre ells els d'avançar cap a un govern obert, fomentar el progrés social, generar valor econòmic...

Els següents en són alguns dels més destacats:

- Permetre la interconnexió entre les administracions (això permetria un intercanvi de dades més eficient, a més d'una dosificació de l'esforç que ha de fer cada organització per extreure dades i poder-les reutilitzar. El fet de compartir dades facilita i agilita la tasca.
- Convertir les administracions públiques en organitzacions totalment transparents. Per això, les dades públiques han d'estar a l'abast no només d'empreses sinó d'usuaris aleatoris i tota la ciutadania en general per fer-ne ús, i que cap dada no quedi amagada ni hi hagi dubtes de com s'estan gestionant certs recursos per part de les administracions.
- Fomentar la participació ciutadana. Es busca que hi hagi una comunicació activa entre ciutadà i govern, que es pugin tirar endavant iniciatives que interessin als ciutadans, etc.
- Conscienciar la societat sobre la importància i el potencial de la reutilització de dades. Canviar la mentalitat de la gent i ensenyar-los que es poden aconseguir moltes coses fent les dades públiques, reutilitzant-les, i intercanviant-les.
- Generar valor econòmic. La reutilització de dades ha de servir a les empreses per a crear nous serveis, i per obrir nous llocs de treball. A més a més, una coordinació amb les administracions farà possible que totes aquestes iniciatives que es creïn tinguin una repercussió social i econòmica.

3.4. Beneficis

Els beneficis de l'Open Data són molts, i poden tenir repercussió a molts nivells de la societat.

- **Beneficis econòmics:** les dades obertes donen la possibilitat a empreses a crear serveis o fins i tot aplicacions que donin informació als ciutadans, de diferents àmbits (transports, cultura, meteorologia, etc), d'aquesta manera obrint un nou mercat.

Dins d'aquest punt és interessant destacar l'estudi fet l'any 2012 per el "*Observatorio Nacional de las Telecomunicaciones y la Sociedad de la Información (ONTSI)*", anomenat "*Estudio de caracterización del sector infomediario*" (Observatorio Nacional de las Telecomunicaciones y la Sociedad de la Información, 2012) (és a dir, del sector de la reutilització de dades), del qual en van sortir uns resultats sorprenents, que reflecteixen que en aquell any la reutilització d'informació pública en el territori espanyol va generar un volum de negoci anual d'entre 300 i 600 milions d'euros, a més a més d'obrir gairebé 4000 llocs de treball per a labors de reutilització de dades obertes. De fet, gairebé un 50% de l'activitat de les empreses considerades en aquest estudi era enfocat a la reutilització de dades. (Iprocuratio, 2010)

Un catàleg ampli de dades obertes, amb els seus respectius estudis per a la reutilització, pot donar lloc a idees d'innovació que en un futur es poden convertir en projectes, serveis, aplicacions o fins i tot productes.

- **Beneficis per a la ciutadania:** per al ciutadà, l'avantatge principal que proporciona l'Open Data és el d'acostar-lo a allò que es coneix com a "Govern Obert", aquell govern capaç d'entaular conversacions amb els ciutadans, de facilitar la seva participació en decisions relacionades amb les administracions públiques, i allò que és més important: aconseguint una transparència de cara al ciutadà pel que fa a la gestió de recursos, serveis...

A més a més, la utilització de dades públiques genera la possibilitat de crear aplicacions i serveis dels quals els ciutadans se'n podran beneficiar.

- **Beneficis per a les administracions públiques:** un punt a favor amb el que s'han trobat les administracions públiques és que anteriorment si es volia invertir en crear aplicacions o serveis de caràcter infomediari havien de comptar amb uns costos i unes despeses exageradament alts. Tanmateix, ara les empreses infomediàries se n'encarreguen, a més generant un valor econòmic de la pròpia reutilització de dades. A part, aquí també es pot establir un lligam amb el segon punt, ja que les mateixes administracions es beneficien d'una elevada participació i col·laboració ciutadana, que poden ajudar a millorar el servei públic amb continguts o idees que han aportat ells mateixos.

El fet que les dades siguin obertes implica que les Administracions Públiques que mostrin aquest tipus de dades tindran una visió més considerada per part del públic, ja que estaran demostrant transparència i responsabilitat respecte els processos que es porten a terme dins la mateixa organització.

Per últim, l'existència de dades obertes facilita l'intercanvi de dades per part de les administracions. D'aquesta manera es disposa d'unes administracions públiques més eficients, els treballadors de les quals no tenen unes càrregues de treball tant

elevades ja que poden col·laborar amb altres administracions.

- **Altres:** en l'àmbit universitari Open Data té un paper fonamental en la integració de dades, en casos concrets ha ajudat a formar una Web sobre Dades Universitàries. És el cas de *Linked Universities* (<http://linkeduniversities.org/>), aliança que dóna suport a aquesta iniciativa i en la qual la UPF n'és un dels membres, publicant les seves dades obertes com a *Linked Data* (o dades enllaçades). (Universidad Pompeu Fabra, 2012).

4. Estat de l'art sobre la disponibilitat i ús de dades

En aquest apartat s'exposarà la recerca feta tan a àmbit nacional com internacional sobre els portals d'Open Data més importants que existeixen, quines temàtiques tracten les dades que exposen, així com els seus formats i estructures.

4.1. Marc tecnològic

4.1.1. Estructura i temàtiques

Com a primer punt, és important destacar que, avui dia, la majoria d'empreses i institucions que fan ús de dades en format obert ho fan a través de pàgines Web.

Per tant, s'ha considerat rellevant estudiar els principals portals que ofereixen dades obertes i observar quines característiques presenten en comú pel que fa tant a la temàtica com a l'estructura de la informació.

El primer que crida l'atenció és que gairebé el 100% dels portals visitats disposa d'una secció introductòria on explica els conceptes principals de l'Open Data i posa èmfasi en la rellevància que té Open Data i els avantatges i beneficis que aporta. Òbviament, dediquen un apartat al catàleg de dades (normalment estructurades en categories que es mencionaran més endavant), especificant els formats en les que estan disponibles així com les condicions legals per a la seva reutilització. La majoria dediquen un apartat a explicar el projecte o projectes que la institució que ha creat el portal està duent a terme mitjançant l'obertura de dades. Normalment afegeixen un apartat amb pàgines webs que s'han fet a partir de la reutilització de dades del seu portal, o també aplicacions que mostren usos diversos d'aquestes dades. Per acabar, habitualment hi ha una secció bé per contactar amb la iniciativa o bé per col·laborar-hi.

Malgrat que, tal i com s'ha comentat en l'apartat anterior, la majoria de pàgines webs presenten estructures similars, hi ha diversos organismes que estan treballant per unificar o estandarditzar el format d'aquestes webs. Un dels projectes més satisfactoris en aquest sentit és l'anomenat CKAN (Comprehensive Knowledge Archive Network). Es tracta d'una aplicació web de codi obert creada per la Open Knowledge Foundation (organització sense ànim de lucre que dona suport a les iniciatives de recursos oberts). Aquesta aplicació no només emmagatzema dades sinó que facilita l'obtenció, la distribució i la utilització d'aquestes, i no està sotmesa a cap tipus de drets de llicència ni a cap restricció (Open Knowledge Foundation, 2004) Actualment hi ha molts portals basats en el CKAN, a més a més, governs importants com l'australià o el del Regne Unit l'utilitzen per a la realització dels

seus catàlegs de dades obertes.

A l'àmbit espanyol es podria destacar la fundació CTIC (Centre de Tecnologies de la Informació i la Comunicació). Aquesta fundació té la tasca de desenvolupar estratègies per millorar la transparència del sector públic mitjançant la publicació de dades en formats oberts i accessibles. Destaca per comptar amb una metodologia pròpia per desenvolupar estratègies d'obertura de dades en Administracions Públiques que ja ha donat resultats satisfactoris, no només nacionalment sinó internacionalment (Fundación CTIC, 2003)

El següent punt important d'anàlisi dels portals Open Data és el de la caracterització de les dades que ofereixen, és a dir, en quines temàtiques s'agrupen. Malgrat hi ha variacions entre els portals s'ha fet un recull de les temàtiques més habituals:

- **Geografia:** s'exposen dades tant de la geografia com de la cartografia del país respectiu.
- **Demografia:** aquest apartat engloba dades demogràfiques i de territori
- **Transport:** trànsits, radars, transport públic...
- **Cultura, oci:** es publica informació d'esdeveniments culturals o de destinacions turístiques, etc...
- **Salut:** dades sanitàries, d'hospitals...
- **Govern i administracions públiques:** subvencions, pressupostos...
- **Educació:** indicadors estadístics sobre els nivells d'educació nacional, centres...
- **Negocis:** ofertes o calendaris de treball
- **Medi ambient:** indicadors mediambientals, residus...

4.1.2. Formats

Potser el punt més important és el de descriure els diferents formats en els quals estan presentades les dades. Al cap i a la fi, és el que determinarà l'ús que se'n faci i la facilitat per treballar-les.

Formats oberts:

- **JSON:** format d'arxiu molt senzill d'interpretar per qualsevol llenguatge de programació i en comparació a altres fàcil de processar per les computadores.

- **XML:** format altament utilitzat per l'intercanvi de dades. Els seus trets característics són la construcció dels arxius, el fet que permet mantenir l'estructura de les dades i que permet als programadors escriure parts de la documentació amb les dades sense risc d'interferir amb ells.
- **CSV:** són els arxius separats per comes. El seu format és compacte i per tant permet la transferència de grans conjunts de dades amb la mateixa estructura. El seu inconvenient és que sense documentació les dades contingudes en aquest format són pràcticament inservibles ja que s'hauria de, gairebé, endevinar el significat de les columnes.
- **RDF:** és una infraestructura que serveix per la representació de recursos de la web. Ho fa en expressions que tenen la següent estructura: subjecte-predicat-objecte. El recurs representat és el subjecte, el predicat és la propietat sobre la qual es pretén establir el recurs, i l'objecte és el valor de la propietat amb la que s'estableix la relació. És un format, però, que per si sol no és interpretable; és a dir que per poder-lo interpretar es farà ús d'un altre format, en concret el XML
- **ODF:** o Open Document. Format estàndard per emmagatzemar dades ofimàtiques, ja siguin gràfiques, presentacions, fulles de càlcul...

Formats propietaris:

- **XLS:** format propietari de Microsoft, pertanyent al programa del paquet Office, Microsoft Excel. Les dades són representades mitjançant files i columnes

Altres formats:

- **ZIP:** és el format de compressió per excel·lència. Serveix tant per a programes lliures com propietaris.
- **FORMATS D'IMATGE:** els més habituals són el JPEG i el PNG.
- **PDF:** format multiplataforma. Document portàtil universal, amb la característica que manté l'aparença del document inicial. Aquest format inicialment era propietari, però des de 2008 no només no ho és, sinó que a més és l'estàndard ISO per a fitxers que continguin documents electrònics per a la seva preservació de llarga durada.
- **SPARQL:** és un llenguatge estandarditzat que serveix per consultar dades en format RDF.
- **API - Web services:** API són les sigles d'interfície de programació d'aplicacions.

Aquesta interfície especifica com haurien d'interaccionar diferents components de programes informàtics. Defineix el "contracte" d'un component informàtic amb qui faci ús dels seus serveis.

4.1.3. Tim Berners-Lee i les seves 5 estrelles

Tim Berners-Lee és un enginyer informàtic britànic i professor del MIT. La fita per la qual és reconegut mundialment és la coordinació del grup d'investigadors del CERN que van desenvolupar el projecte *World Wide Web* (WWW). Va ser ell qui va implementar la primera comunicació entre un client i un servidor HTTP a través d'internet. A més a més, és el director del *World Wide Web Consortium* (W3C), que supervisa i estandarditza el desenvolupament de la web i les tecnologies sobre les quals es fomenta.

En relació a l'explotació de les dades, Berners-Lee va crear un sistema de classificació per establir quina qualitat tenen les dades i quina facilitat per processar-les presenten (Berners-Lee, W3C, 1994). Aquesta classificació és anomenada la de "les 5 estrelles", i funciona de la següent manera:

QUALIFICACIÓ	FORMAT I TECNOLOGIA USATS
★	Format no estructurat, llicència oberta
★ ★	Format estructurat, però propietari
★ ★ ★	Format estructurat i obert
★ ★ ★ ★	Utilització de URLs per identificar i referenciar les dades
★ ★ ★ ★ ★	Dades importants enllaçades amb altres de manera que s'estableix un context (el que es coneix com Linked Data)

Taula 1. Les 5 estrelles de Berners-Lee per a la classificació de dades

Una estrella s'aconsegueix en el cas que les dades siguin de tipus públic i amb llicència oberta. Normalment serà fàcil visualitzar les dades, guardar-les o imprimir-les (ja que els formats englobats en aquest grup ho permeten) però la dificultat vindrà a l'hora de reutilitzar-les. Exemples d'aquest esglaió més baix de la classificació són formats d'escriptura com el PDF, que dificulta l'extracció de les dades, ja que és difícil que conservin el seu format; formats d'imatge (jpeg, png), formats de música (mp3...), o de vídeo, alguns dels més

comuns son avi, mp4, o mpg.

Per a que siguin classificades amb dos estrelles, a més dels punts comentats anteriorment, les dades han d'haver estat publicades en algun format estructurat. El clar exemple són els arxius d'Excel, que sí que permeten una anàlisi i un tractament més sistemàtic de les dades (a més de permetre'n l'exportació a altres formats), però en ser un format propietari requereix llicències no públiques per accedir-hi.

La diferència entre dues estrelles i tres estrelles és el fet d'afegir la característica de no propietat al format estructurat. Fent la comparativa amb l'Excel, un format XLS tindria 2 estrelles mentre que utilitzar un CSV implicaria una classificació de 3 estrelles. Utilitzar un format no propietari habilita a l'usuari l'extracció de dades sense cap tipus de limitació de software, a més a més s'estalvien els costos de llicències.

Quatre estrelles se li atorguen a aquells conjunts de dades que fan ús de URLs per identificar-les, de manera que les dades van adquirint un significat en funció de les etiquetes utilitzades a l'hora de realitzar el document. Això és òptim quant a utilització i transformació de dades, tanmateix és una tasca costosa per a qui publica les dades, no només econòmicament sinó en hores de dedicació, ja que s'hauran de separar totes les dades i anar assignant les URLs que els corresponguin.

La cinquena estrella s'obté quan, a més de complir els punts de l'apartat anterior, s'uneixen algunes de les dades més importants amb dades pertanyents a altres fonts o persones, creant d'aquesta manera una mena de xarxa d'informació. Dins d'aquest grup, el format per excel·lència és el citat anteriorment RDF.

4.2. Llistat de recursos en l'àmbit espanyol



Figura 1. Distribució de les iniciatives Open Data a Espanya. Font: CTIC

Aquest apartat està enfocat a la recerca d'iniciatives i portals que ofereixen catàlegs amb dades obertes dins d'Espanya. Alguns dels que es comentaran van ésser creats per la iniciativa Open Data. En el mapa adjunt (Figura 1) es pot veure la seva distribució, ja siguin d'àmbit nacional, autonòmic o local.

Altres, però, agrupen temes més concrets i no han estat creades per a Open Data, sinó que provenen d'entitats amb molta història. Val a dir que dins d'aquest grup es troben entitats que també subministren a portals Open Data.

Tornant al mapa, s'hi pot veure que actualment existeixen una gran quantitat d'iniciatives repartides pel territori espanyol. Per fer-ne una caracterització és interessant classificar-les segons els continguts que presenten, és a dir si són d'àmbit nacional, autonòmic o local.

4.2.1. Nacional

- **datos.gob.es/**

A àmbit nacional, o estatal, existeix el portal del *Gobierno de España*, **datos.gob.es**, l'encarregat de gestionar el *Catálogo de Información Pública* de les Administracions Públiques.

Aquest portal va néixer l'any 2011 sota el marc d'accions del projecte Aporta. Pertany al *Ministerio de Industria, Energía y Turismo* i al *Ministerio de Hacienda y Administraciones Públicas* del Govern Espanyol, i hi col·labora el CTIC.

L'objectiu principal de la plataforma és el de fomentar l'obertura i la reutilització de la informació del sector públic nacional i conscienciar la societat sobre la rellevància i el valor que té la cultura de dades obertes. És per això que s'hi poden trobar dades el sector públic de moltes categories diferents, entre elles: economia, turisme, educació... (Ministerio de Industria, Energía y Turismo y Ministerio de Hacienda y Administraciones Públicas, 2011)

A més del catàleg de dades, és particularment interessant el fet que s'ofereixen recursos divulgatius i fins i tot pràctics per poder desenvolupar serveis que es basin en la reutilització de dades.



Figura 2. Pàgina principal portal datos.gob.es/

4.2.2. Autonòmic

Pel que fa a àmbit autonòmic, a la taula adjunta (taula 2) es mencionaran les iniciatives de les diferents Comunitats Autònomes:

COMUNITAT AUTÒNOMA	PORTAL OPEN DATA
Andalusia	juntadeandalucia.es/datosabiertos
Aragó	opendata.aragon.es
Castella - La Manxa	transparencia.castillalamancha.es/opendata
Castella i Lleó	datosabiertos.jcyl.es
Catalunya	dadesobertes.gencat.cat
Galícia	abertos.xunta.es
Illes Balears	caib.es
Illes Canàries	opendatacanarias.es
La Rioja	larioja.org
Navarra	gobiernoabierto.navarra.es
País Basc	opendata.euskadi.net

Taula 2. Portals Open Data de les Comunitats Autònomes Espanyoles. Font: CTIC

Per entendre les característiques d'aquest tipus de portals s'ha decidit analitzar el de Catalunya, el portal de dades obertes del **gencat**.

- **dadesobertes.gencat.cat**

Aquest portal el gestiona la Direcció General d'Atenció Ciutadana, i s'hi publiquen dades de molts dels organismes públics de la Generalitat. La direcció legal del portal va ser a càrrec de la Direcció General d'Atenció Ciutadana (Generalitat de Catalunya, 2011), però la direcció tècnica va ser cosa del CTIC.

Com a característica important, destacar que les dades que presenta són majoritàriament en formats estructurats (si bé alguns propietaris). I com a novetat, permet la descripció dels catàlegs en format RDF.



Figura 3. Pàgina principal portal <http://dadesobertes.gencat.cat/ca/>

4.2.3. Local

A àmbit local els portals d'Open Data tenen estructures molt similars als d'àmbit autonòmic, no només per com es troba estructurat el seu contingut sinó també per la temàtica de les dades que ofereix. Això sí, la seva informació prové exclusivament de dades del respectiu municipi o ciutat.

En la següent taula s'esmenten algunes de les ciutats que ofereixen portals d'Open Data i les seves adreces electròniques.

CIUTAT	PORTAL OPEN DATA
Badalona	badalona.cat
Barcelona	opendata.bcn.cat
Bilbao	bilbao.net
Madrid	datos.madrid.es
Màlaga	datosabiertos.malaga.eu
Manresa	bigovopendata.bismart.com
Pamplona	pamplona.net
Rubí	opendata.rubi.cat
Sabadell	sabadell.cat
San Sebastian	donostia.org
Santander	datos.santander.es
Saragossa	zaragoza.es
Tarragona	opendata.tarragona.cat
Terrassa	opendata.terrassa.cat
València	valencia.es

Taula 3. Portals Open Data de diferents ciutats Espanyoles. Font: CTIC

4.3. Llistat de recursos en l'àmbit internacional



Figura 4. Distribució de les iniciatives Open Data al món. Font: CTIC

El mapa adjunt (Figura 4) mostra la distribució d'iniciatives Open Data arreu del món. A data d'avui hi ha contats més de 250 portals repartits en 40 països (la major part a Europa i Amèrica del Nord).

A continuació es comenten alguna de les iniciatives més influents. En relació a això és important destacar que la World Wide Web Foundation's Open Data Index (World Wide Web Foundation, 2012) ha establert un ordre per classificar quin són els països capdavaners en aquest aspecte. Alguns dels que apareixen en les primers posicions són: els ja previsibles Estats Units, Regne Unit, Noruega, i d'altres més sorprenents com Nova Zelanda o Corea.

- **data.worldbank.org**

Aquest portal està gestionat pel Banc Mundial. Aquesta és una institució financera de les Nacions Unides que dona préstecs a països en vies de desenvolupament amb l'objectiu primordial de la reducció de la pobresa. És un membre del *World Bank Group* i un membre del *United Nations Development Group*.

El portal és de caire internacional i posseeix un catàleg molt gran de dades a nivell global. Les dades que s'hi poden trobar tenen a veure amb els índexs de desenvolupament. S'hi poden trobar més d'un miler d'indicadors, alguns amb dades

recopilades des de fa dècades, i milers i milers d'altres col·leccions d'estadístiques sobre desenvolupament, educació...

- **data.gov**

Aquest és el catàleg de dades públic i d'àmbit nacional dels Estats Units. No només inclou un elevat nombre de departaments de l'administració pública sinó també moltes entitats públiques. Des del mateix any de la seva creació (el 2009) es va establir una directiva que obligava a totes aquelles agències que tinguessin a veure amb l'administració pública dels Estats Units a publicar mínim 3 datasets d'un valor elevat en un període de 45 dies.

Avui dia és considerat com un dels portals més importants a nivell mundial, no només per la gran quantitat de dades que s'hi poden trobar sinó també per la seva qualitat.

Les dades que presenta es troben en formats estructurats, compleixen els requisits per a la reutilització i s'hi poden trobar fins a 20 categories diferents, entre elles seguretat pública, comerç, salut, finances...

- **data.gov.uk**

Portal dedicat a la reutilització de la informació pertanyent al Govern britànic. La informació que aporta és de caràcter públic i és d'àmbit nacional britànic. S'hi pot trobar informació no només del govern central sinó també dels regionals i locals. Com a punt destacat, esmentar que utilitza el software CKAN (veure apartat anterior 5.1.1).

Es considera un dels portals més complets degut a la distribució de la informació que presenta, amb enllaços de contacte amb l'editor i amb recursos que tinguin a veure amb la informació presentada, totes les dades venen acompanyades d'una descripció feta pels usuaris, amb la possibilitat de puntuar-les, de cercar-les fàcilment, etc. A més a més inclou la possibilitat d'utilització del SPARQL (veure apartat 4.1.2 sobre formats), per explotar les dades de forma automatitzada.

- **datos.gob.mx**

Aquest portal és l'equivalent als dos esmentats anteriors però del govern mexicà. És el lloc oficial de dades obertes del govern de la República de Mèxic, i s'hi pot trobar el catàleg de dades obertes que eventualment es pretén que compregui totes les dades de l'administració pública mexicana. És considerat capdavanter en comparació a altres països, segons l'índex mundial esmentat anteriorment.

Una característica interessant del portal és que incorpora un apartat anomenat

"*Historias*" on s'explica a l'usuari diferents narratives i exemples que el poden ajudar a entendre quina utilitat té la reutilització de dades, quins projectes s'estan assolint a nivell de desenvolupament social, econòmic i polític a Mèxic.

En el portal s'hi poden trobar centenars de conjunts de dades de diverses institucions i de temes variats com el comerç, les exportacions, investigacions...

- **data.norge.no**

Noruega, actualment, en la última versió del World Wide Web Index (la de l'11 de desembre de 2014), encapçala l'índex i és considerat el país que més benefici social, econòmic i polític obté de la Web.

La seva web, donant validesa a la classificació del ranking, és una de les més completes, i conté datasets compresos en 22 categories.

- **Freebase.**

Freebase és un portal col·laboratiu de coneixement que conté multitud de dades bàsicament aportades pels seus membres, que són tant professionals com voluntaris. Va ser desenvolupat per la companyia americana *Metaweb*, i va ser adquirida al 2010 per *Google*. El seu objectiu és crear una comunitat global que permeti accedir a la informació de manera més eficient. El llistat d'entitats col·laborades fins a dia d'avui supera els 20 milions.

- **Wikidata.**

Wikidata és una base de dades col·laborativa, operada per la fundació Wikimedia. El seu propòsit és el de crear una font comuna de certs tipus de dades que pugui ser usada per altres projectes de Wikimedia com la Wikipedia.

- **YAGO**

Base de dades desenvolupada al Max Planck Institute for Computer Science in Saarbrücken. A data de 2012 YAGO comprenia coneixement de més de 10 milions d'entitats i més de 120 milions de fets sobre aquestes entitats. La informació recopilada és extreta de fonts com Wikipedia, WordNet, GeoNames...

4.4. Portals que exploten les dades de manera visual i interactiva

Posseir quantitats molt grans de dades obertes està bé, però un es pot trobar amb dificultats per treballar-les o fins i tot per entendre-les. De vegades el fet de recol·lectar moltes dades no garanteix un avenç en el tema d'estudi, ja que les dades per si soles poden no aportar la informació desitjada. És en aquest punt on entra en joc la representació adequada d'aquestes dades.

A partir d'aquesta necessitat de representar les dades recol·lectades, cada cop estan naixent més portals que no només exploten les dades sinó que faciliten el següent punt: les representen de manera visual. És a dir, ajuden a l'usuari a comprendre quin significat tenen aquestes dades recol·lectades, quina interconnexió tenen entre elles, quines tendències segueixen les dades, com han evolucionat al llarg del temps, en resum: quines conclusions es poden extreure de les dades recol·lectades.

No hi ha un estàndard per a la representació de dades, sinó que de vegades serà útil un gràfic en forma de formatget, altres un gràfic de barres, mapes interactius, gràfics amb eixos intercanviables, etc.

A continuació es presentarà un llistat d'iniciatives i portals que exploten les dades i les representen de manera interactiva promovent que l'usuari estàndard pugui entendre de manera senzilla allò que està veient:

- **Gapminder.org**

La fundació Gapminder és una organització sense ànim de lucre creada a Stockholm, Suècia. El seu principal objectiu és el de promoure objectius sostenibles de desenvolupament i de compliment dels *United Nations Millennium Development Goals* mitjançant l'ús i la comprensió d'estadístiques i altres indicadors informatius relacionats amb el desenvolupament social, econòmic i mediambiental a nivell no només local i nacional sinó també global.

-

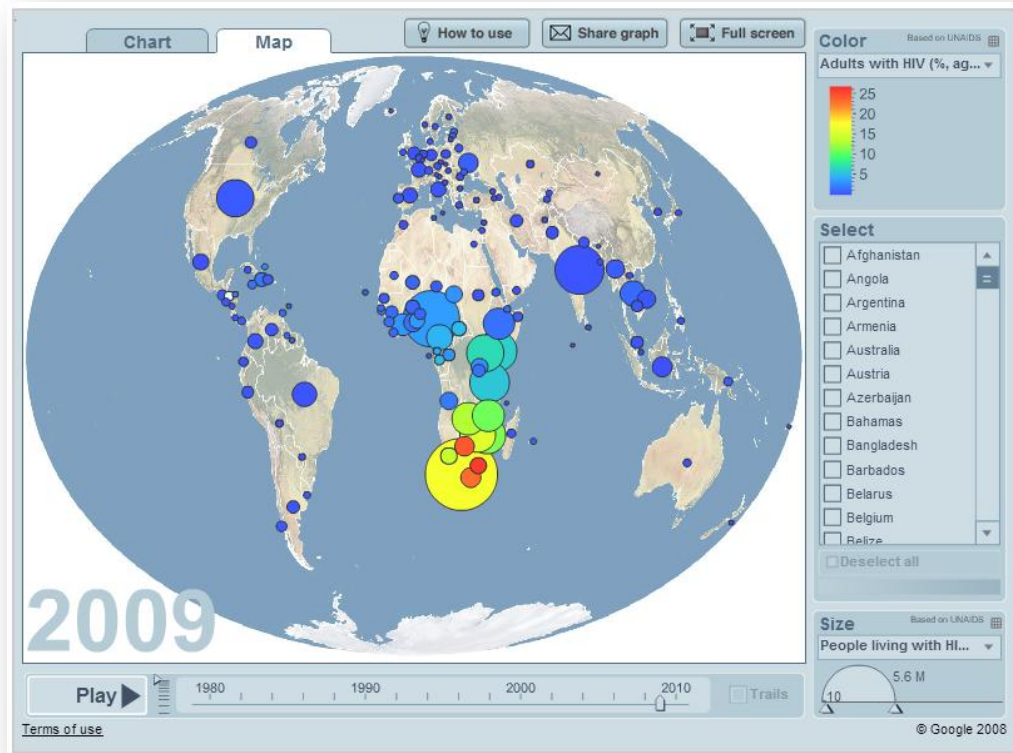


Figura 5. Exemple de gràfic de Gapminder. % d'adults amb HIV per país i per any

- **Eixos.cat**

EIXOS és un observatori econòmic digital creat per l'empresa Planol.info a Barcelona l'any 2011, que analitza i fa seguiment de l'activitat econòmica de les ciutats i els polígons industrials per a l'elaboració d'estudis de mercat. Consta d'una plataforma digital web i una web mòbil. Classifica i cataloga comerços urbans, empreses, naus o locals buits, edificis administratius, transport públic... Per tant emprenedors, franquiciats, cadenes d'establiments, inversors, administracions públiques o investigadors són els usuaris potencials d'EIXOS.

A EIXOS es combinen dades provinents de treball de camp, recollides amb mitjans propis periòdicament, amb dades obertes provinents de les administracions públiques i d'altres entitats privades.

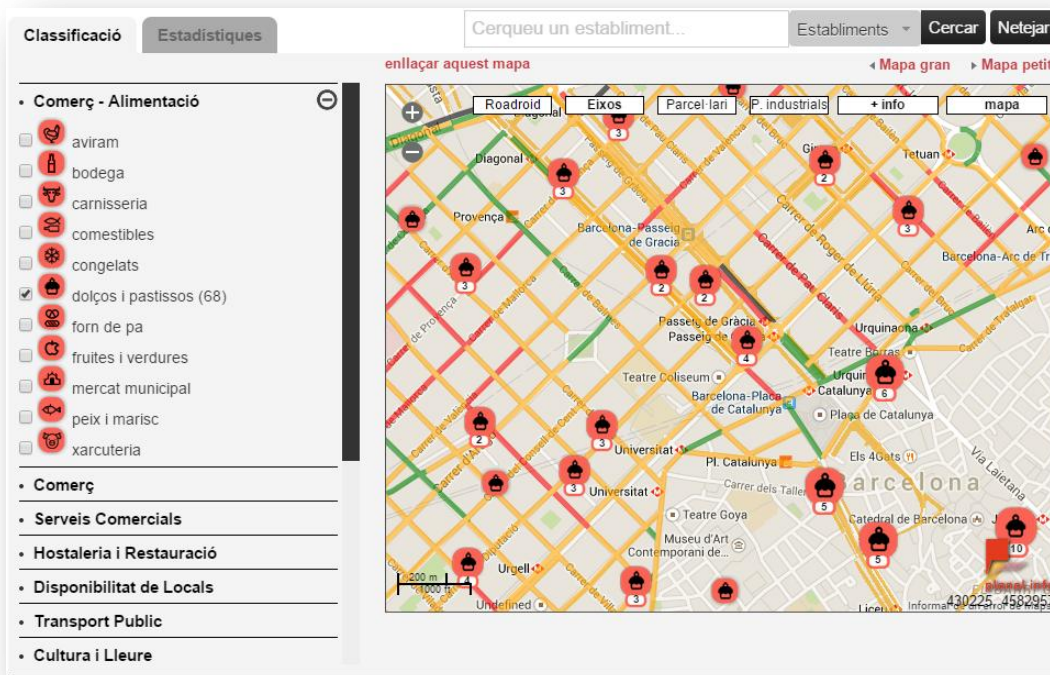


Figura 6. Exemple de mapa visual a Eixos.cat. Distribució de pastisseries i botigues de dolços

- **atlas.media.mit.edu**

L'Observatori de Complexitat Econòmica és una eina de dades de comerç internacional. El seu objectiu és el de distribuir dades d'exportacions entre països, i els productes que exporten, de manera visual.

El projecte va ser dut a terme al MIT, concretament al *MIT Media Lab Macro Connections Group*.

Actualment l'observatori compren més de 20 milions de visualitzacions interactives, connectant així centenars de països a les seves destinacions d'exportació i amb els productes amb els que comercien.

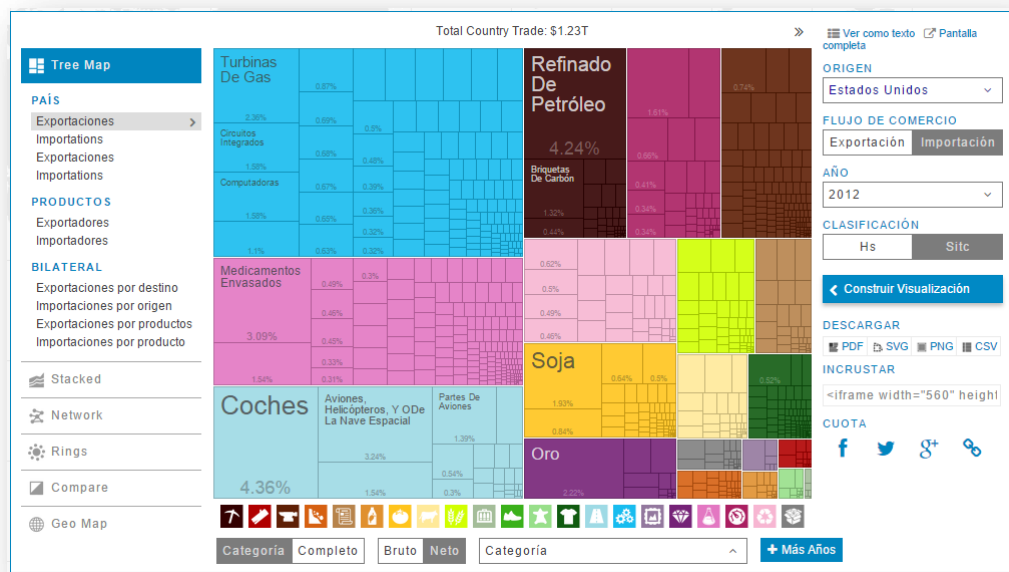


Figura 7. Exemple de visualització de l'Observatori de Complexitat Econòmica. Productes exportats per USA el 2012

- **InstantAtlas.**

InstantAtlas és una eina de visualització de dades i d'informes utilitzada per a la comunicació de dades estadístiques basades primordialment en la localització. Aquesta eina va ser creada per l'empresa Geowise Ltd, una companyia de software fundada el 1997 i amb seu a Edinburgh.

InstantAtlas és utilitzada per analistes, investigadors i estadístics. Conté plantilles predefinides on carregar les dades, amb infinitat de varietats per adaptar-se a les necessitats de l'usuari.

La utilitat dels seus informes dinàmics és la de presentar les dades de manera gràfica i intuïtiva per permetre a l'usuari descobrir tendències, relacions entre dades patrons i fins i tot anomalies.

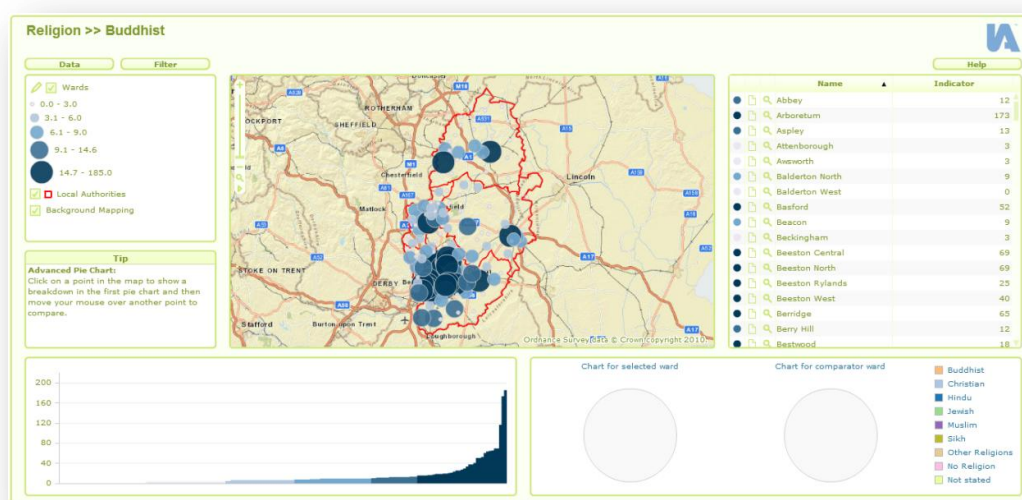


Figura 8. Exemple d'InstantAtlas. Distribució de budistes en poblacions del Regne Unit.

- **Visual.ly**

Visual.ly consisteix en una plataforma comunitària per a la visualització de dades. Va ser fundada el 2011 per Stew Langille, Lee Sherman, Tal Siach, i Adam Breckler. Conté una sèrie d'eines que permeten, de manera senzilla, crear unes representacions bastant espectaculars, a més a més d'una plataforma per compartir totes les creacions.

A part de la gran utilitat com a eina, destacar que Visual.ly també serveix com a comunitat per a persones que es dediquen a la recerca o fins i tot dissenyadors, i molts usuaris vinculen les seves comptes amb les xarxes socials.

El portal permet als usuaris cercar imatges mitjançant etiquetes, descripcions o fonts, i les seves dades són variades, tocant sectors com la política, l'economia o el comerç.

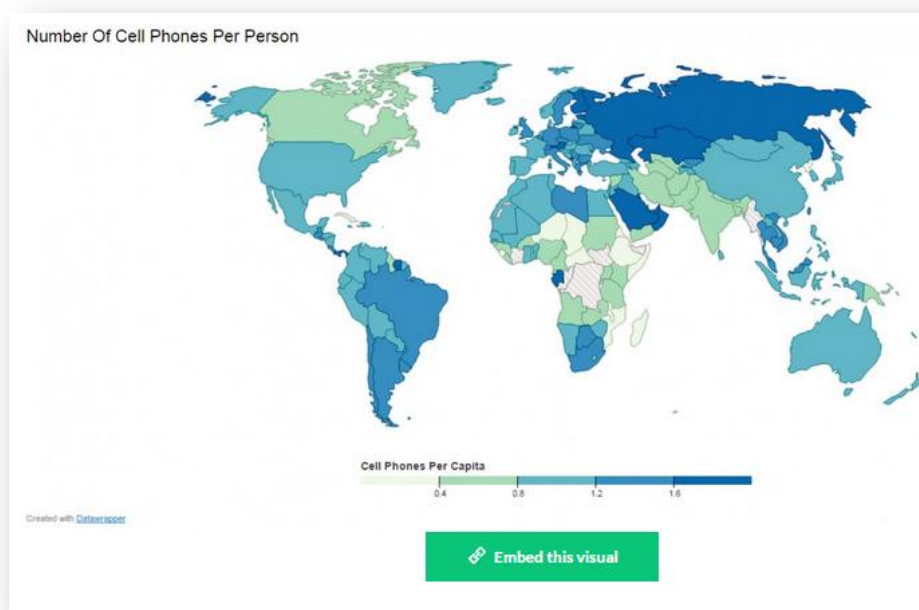


Figura 9. Exemple de visualització de Visual.ly. Telèfons mòbil per persona i per país.

5. Estudi de les opcions per fer una interfície d'explotació de dades

Un cop estudiats els llocs d'on es poden treure dades obertes, i haver fet una cerca sobre com presentar aquestes dades d'una forma visual i fàcil d'interpretar, es considera que hom és capaç de presentar una explotació de dades personalitzada.

El primer que caldrà valorar és amb quin format es podria explotar adequadament una sèrie de dades de manera que es puguin fer gràfiques, mapes, taules, etc. Per aconseguir això s'han escollit tres plataformes que podrien ser útils per aquesta tasca i se n'han valorat els pros i els contres, per posteriorment triar el que resulti més adequat.

A continuació s'expliquen les possibilitats escollides, els seus avantatges i inconvenients:

5.1. Python



Icona:

Lloc Web: <https://www.python.org/>

La primera opció estudiada ha estat la de programació amb Python.

Python és un llenguatge de programació altament usat en l'actualitat. La seva creació va esdevenir-se a càrrec de Guido van Rossum (programador informàtic holandès actualment treballant a Google) l'any 1991.

El que persegueix aquest llenguatge és fer més llegible el codi i alhora permetre elaborar codis més concisos, que no requereixin tantes línies per programar la mateixa comanda que en altres llenguatges.

Pel que fa als propòsits del projecte, la programació amb Python permetria a l'usuari interactuar amb les dades escollides, creant gràfics mitjançant l'ús de les comandes del llenguatge adequades.

Com a punt a favor del llenguatge Python es troba el fet que és completament gratuït. Tanmateix un s'adona que per a la tasca escollida el procés és laboriós, i és difícil obtenir els resultats desitjats si només es tenen nocions bàsiques d'aquest llenguatge.

És cert que amb una certa preparació es podrien crear fitxers amb funcions estadístiques i matemàtiques que serien capaces d'elaborar gràfiques que portarien a la interpretació correcta de les dades estudiades. Això sí, els resultats mancarien d'espectacularitat (visualment parlant), i per tant quedarien pobres respecte altres opcions.

```
# Python 3: Fibonacci series up to n
>>> def fib(n):
>>>     a, b = 0, 1
>>>     while a < n:
>>>         print(a, end=' ')
>>>         a, b = b, a+b
>>>     print()
>>> fib(1000)
0 1 1 2 3 5 8 13 21 34 55 89 144 233 377 610 987
```

Figura 10. Exemple de Codi programat amb Python. Sèries de Fibonacci. Font: <https://www.python.org/>

5.2. Flash



Icona: **ADOBE FLASH**

Lloc web: <http://www.adobe.com/es/products/flash.html>

Flash és un programa d'edició multimèdia d'Adobe (fins el 2005 es deia Macromedia Flash) i a més a més un reproductor (Adobe Flash Player) usat per executar els arxius de tipus Flash.

A dia d'avui Adobe Flash és un dels programes més usats pels dissenyadors, desenvolupadors i animadors tant de pàgines webs com aplicacions mòbil.

Aquest programa d'edició utilitza imatges de mapa de bits, so codi, i vídeo així com gràfics vectorials.

Els fitxers Flash normalment adopten l'extensió SWF i es troben en moltes pàgines web.

Flash també ha estat seleccionada com una opció per a presentar de manera gràfica les dades explotades, però un cop més ens trobem amb el fet que és un procés massa difícil per un usuari ocasional. És cert que els resultats que es podrien obtenir segurament serien molt atractius pel públic, i incorporarien la capacitat de no només limitar-se a gràfiques sinó que el contingut es podria animar. Tanmateix el programa no és idoni per al tractament d'arxius de dades, sinó que és més enfocat al disseny. A més a més, Flash no és gratuït sinó que s'ha de pagar una llicència mensual.

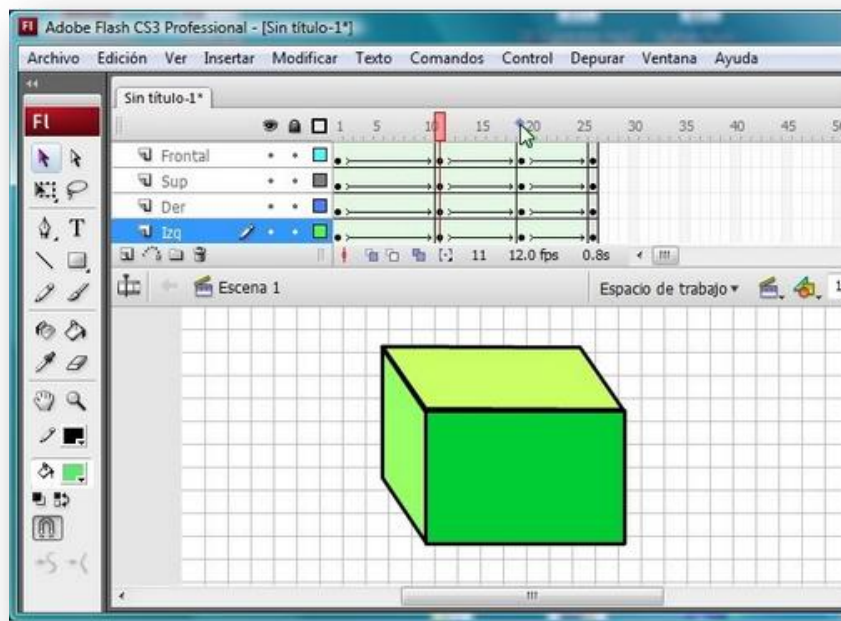


Figura 11. Exemple de disseny usant Flash

5.3. Shiny de RStudio



Icona:

Lloc Web: <http://shiny.rstudio.com/>

Shiny és una plataforma creada per **RStudio**, un medi de desenvolupament, gratis i de codi lliure que es basa en R, el llenguatge de programació per estadística i gràfics per excel·lència.

RStudio està disponible tant per a Microsoft Windows, com per a Mac OS X com per a

Linux. Està programat en codi C++ i utilitza la infraestructura Qt per la seva interfície gràfica.

RStudio va iniciar-se als voltants de desembre de 2010, per tant es pot dir que és un medi novedós, i la seva primera versió BETA pública va ser presentada oficialment el febrer de 2011.

Per entendre el funcionament de RStudio és bàsic conèixer com funciona el llenguatge de programació R, ja que RStudio finalment ve a ser una interfície visual de R que fa més atractiva i més "user-friendly" l'aplicació i l'ús de R a l'usuari. De fet Shiny pot ser instal·lat des del mateix R.

R és un llenguatge de programació, a més a més d'un entorn de desenvolupament de software per a l'obtenció de càlculs i gràfics estadístics. Va aparèixer l'any 1993 i els seus creadors van ser George Ross Ihaka i Robert Gentleman, professors d'estadística de la Universitat d'Auckland, Nova Zelanda (R Project, 2010).

R és el llenguatge més emprat per a l'anàlisi de dades, així com per a desenvolupar programes estadístics. Com a particularitat, destacar que el codi de R és gratuït sota la Llicència Pública General GNU.

El seu funcionament és mitjançant una interfície de comandes. Es pot escriure el codi a la mateixa interfície o en fitxers de funcions, que després serveixen per a obtenir certs resultats. En destaca la facilitat per extreure dades importants de fitxers i la senzillesa per a tractar fitxes amb quantitats immenses de dades. És un llenguatge intuïtiu i que presenta la possibilitat d'incorporar paquets extres que li afegeixen funcions.

Shiny, doncs, funciona com a paquet de R, i és una eina dissenyada per a crear les interfícies visuals més adequades per a la interpretació de dades. A més de ser gratuït, és relativament fàcil fer-lo servir per a un usuari que ja conegui R. A més a més, a la pròpia web <http://shiny.rstudio.com/> s'hi pot trobar un tutorial, així com diverses demostracions de les plataformes que es poden dissenyar.

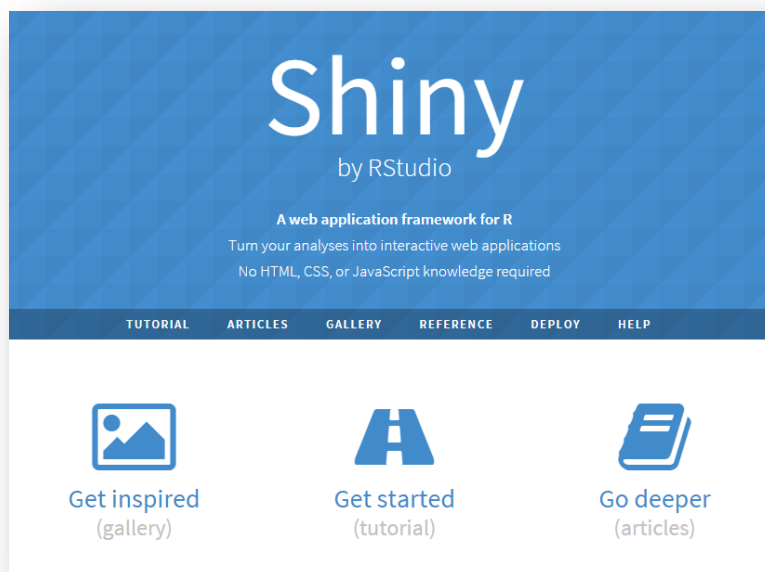


Figura 12. Pàgina Inicial <http://shiny.rstudio.com/>

5.4. Decisió

Finalment, s'ha optat per utilitzar Shiny de RStudio, ja que és l'eina òptima per crear una interfície d'aquest tipus.

A continuació s'inclou una taula comparativa amb els aspectes que han fet decantar-me per la opció de Shiny de RStudio:

	Python	Flash	Shiny
Compatibilitat en diferents formats	3	2	4
Dificultat d'aprenentatge	4	1	3
Gratuït	5	1	5
Idoneïtat per al propòsit escollit	2	2	5
Espectacularitat del resultat	2	5	4

Taula 4: Comparativa entre les 3 opcions per desenvolupar una interfície d'explotació de dades

La taula atorga puntuacions de l'1 al 5 (sent 5 l'òptima i 1 la pitjor) a cada plataforma o programa per cada una de les variables que s'han considerat rellevants. Les puntuacions posades no provenen d'un estàndard oficial sinó que són fruit de l'experiència adquirida a partir de la recerca i l'estudi de les opcions.

Mitjançant un senzill càlcul, pel qual la suma dels valors de les variables serà la puntuació final de cada opció, s'obté:

- Python =16
- Flash =11
- Shiny= 21

Així doncs, l'opció Shiny és la que globalment proporciona l'opció òptima.

El paquet Shiny s'ha dissenyat especialment per al propòsit de la interpretació de conjunts de dades. No només presenta gràfics simples sinó que també permet a l'usuari interactuar amb ells, és a dir canviar els eixos, la cronologia, visualitzar diferents tipus d'interfícies com mapes i gràfics interactius, etc.

Com a punt a favor destacar que és una aplicació gratuïta, que es pot obtenir mitjançant la descàrrega del programa R i instal·lant el paquet corresponent.

És cert que encara no és una eina gaire coneguda però s'ha considerat que no és pas difícil d'usar si es coneix R. A més, la mateixa web facilita la tasca d'aprenentatge amb un tutorial útil, però bàsic, que naturalment s'ha dut a terme, i a sobre conté una galeria amb exemples sobre el seu potencial: mapes interactius, gràfiques amb eixos intercanviables, diagrames de punts, i taules visuals són alguns dels exemples que aporta Shiny.

La majoria d'aquestes gràfiques no es podrien assolir amb Python, requeriria d'una programació molt més elaborada per la qual un usuari aficionat no està preparat, i en cas que s'aconseguissin gràfics adequats seria molt complicat i en alguns casos impossible animar-los. En el cas del Flash si que seria possible obtenir gràfiques d'aquest tipus i potser fins i tot visualment més atractives, ja que és un programa enfocat al disseny, però el fet de necessitar una llicència de pagament, la dificultat del propi disseny, i la dificultat que representaria tractar les dades ha estat un punt clau per decantar-se pel paquet Shiny de R.

Pel que fa al tractament de dades R és un programa capdavanter en aquest àmbit, que permet tractar conjunts de dades independentment de quantes n'hi hagin, i optimitza la cerca de dades encara que hi hagi espais buits.

La gran particularitat de Shiny és que no només podem incorporar tot tipus d'interfícies, sinó que a més l'usuari podrà "jugar" amb elles. És a dir, podrà canviar els eixos, seleccionar diferents rangs, incorporar cronologia i mirar l'evolució d'una gràfica al llarg del temps mitjançant un sol click, obtenir informació a l'instant d'un punt concret d'un mapa de punts, veure dades incorporades en mapes reals, etc.

6. Explotació i representació d'una sèrie de dades

6.1. Obtenció de les dades

Per fer una explotació de dades, òbviament el primer que es necessita és un o més conjunts de dades.

La recerca feta prèviament ha permès tenir una gran varietat de llocs webs d'on es podien recopilar grans quantitats de "datasets" d'infinits de categories diferents, no només d'àmbit nacional sinó també internacional.

Després d'estudiar les diferents possibilitats s'han acabat escollint conjunts de dades pertanyents a la pàgina del Banc Mundial: **data.worldbank.org**.

Val a dir que una de les pàgines més complertes i que proporcionaven dades de més importància pel que fa a la proximitat i la varietat dels temes era la pàgina oficial de l'open data a Barcelona (**opendata.bcn.cat**). A més a més, és molt interessant el fet que gairebé totes les dades estan distribuïdes per districtes, i fins i tot ordenades i separades per cada un dels 73 barris. D'aquesta manera es podia fer una interfície visual de Barcelona que permetia comparar diferents aspectes de la societat i l'economia barcelonina per a cada un dels barris.

Finalment, però, s'ha considerat que la Web del Banc Mundial aportava un salt de qualitat important no només perquè cobreix més temes que altres portals, sinó perquè cronològicament és de llarg el més complet. Per exemple, el portal **opendata.bcn.cat** té l'inconvenient que és un portal jove, i en el millor dels casos s'hi troben dades des de l'actualitat fins a 3 o 4 anys enrere. En canvi en el Banc Mundial, els indicadors més complets contenen dades recol·lectades des dels anys 60. D'aquesta manera, les interfícies visuals que es puguin crear a partir de les dades explotades gaudiran d'una comparació cronològica, que aporta una informació extra, i que en la majoria de casos és rellevant, a més que permet veure l'evolució dels indicadors estudiats.

Com a altre punt important, el Banc Mundial té indicadors de gairebé tots els països del món. Així, no es limita la comparativa a gràfiques, sinó que també es poden crear mapes interactius, que visualment per a l'usuari són molt fàcils d'entendre i permeten comparar la informació amb només un cop d'ull.

El portal **data.worldbank.org** està organitzat de manera que un pot cercar dades tant per país, com per indicador, com per tòpic:

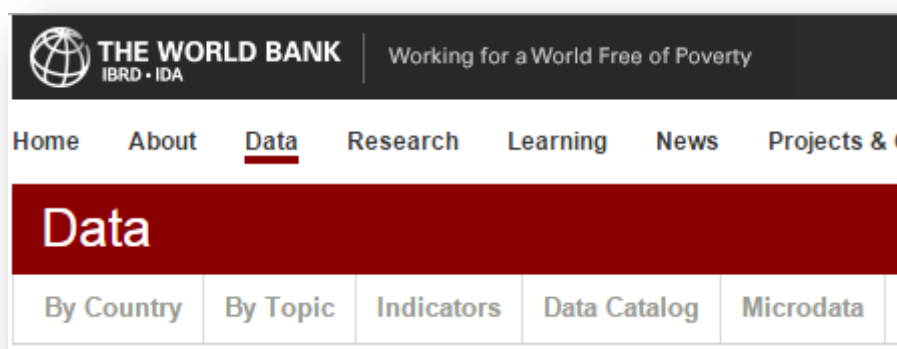


Figura 13. Distribució de conjunts de dades al portal data.worldbank.org

A més a més, conté diversos apartats que resulten molt útils, que expliquen les iniciatives que s'està duent a terme quant a Open Data, els avenços a nivells de xarxes socials, i fins i tot un apartat de contacte.

Fins a 20 temes relacionats amb l'economia, la política, aspectes socials, educació, salut, etc. són els que es poden trobar. A continuació es mostra una imatge dels diferents temes:

Topics	
Agriculture & Rural Development	Health
Aid Effectiveness	Infrastructure
Climate Change	Poverty
Economy & Growth	Private Sector
Education	Public Sector
Energy & Mining	Science & Technology
Environment	Social Development
External Debt	Social Protection & Labor
Financial Sector	Trade
Gender	Urban Development

Figura 14. Temes que agrupen els conjunts de dades de data.worldbank.org

Aquest portal és considerat dels millors en el seu apartat per la seva estructura. No només per la quantitat d'indicadors que conté sinó per la informació que s'hi troba en cada indicador. Agafant l'exemple del producte interior brut (en \$), el primer que un troba és una breu descripció sobre l'indicador:

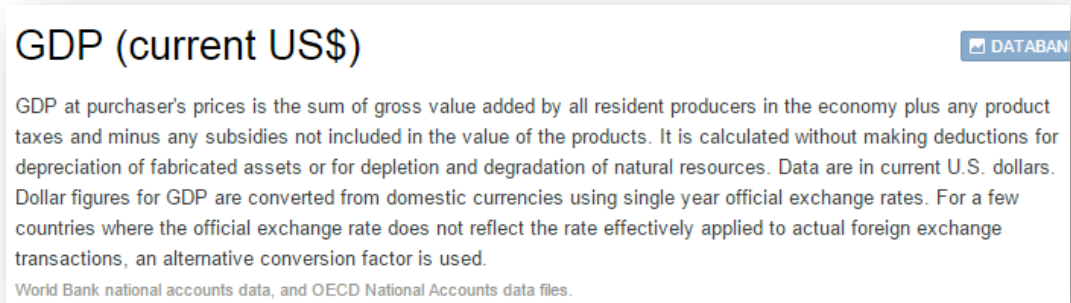


Figura 15. Explicació del càlcul del producte interior brut en el portal data.worldbank.org

El següent punt important és que les dades contingudes es poden descarregar tant en format EXCEL, com CVS, com XML:

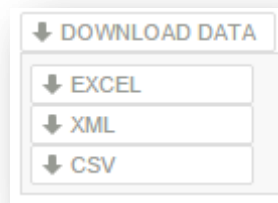


Figura 16. Formats en els que es poden descarregar les dades a data.worldbank.org

Indagant en els diferents apartats, el que s'ha fet és buscar conjunts de dades que presentessin les següents característiques:

- Dades que aportin informació rellevant a nivell econòmic o social
- Dades cronològicament completes fins aproximadament els anys 60
- Dades completes pel que fa a un % elevat dels països del món
- Dades que tinguin certa relació entre elles pel que fa a la temàtica de manera que tinguin sentit relacionar-les entre elles en un mateix gràfic

D'aquesta manera s'han escollit cinc indicadors que complien les característiques mencionades anteriorment:

	A	B	C	D	E	F
1	Country Name	Country Code	Region	IncomeGroup	SpecialNotes	
2	Aruba	ABW	Latin America & Caribbean	High income: nonOECD	SNA data for 2000-2011 are updated from official government statistics	
3	Afghanistan	AFG	South Asia	Low income	Fiscal year end: March 20; reporting period for national accounts data: n	
4	Angola	AGO	Sub-Saharan Africa	Upper middle income	April 2013 database update: Based on IMF data, national accounts data	
5	Albania	ALB	Europe & Central Asia	Upper middle income		
6	Andorra	AND	Europe & Central Asia	High income: nonOECD		
7	Arab World	ARB			Arab World aggregate. Arab World is composed of members of the Lea	
8	United Arab Emirates	ARE	Middle East & North Africa	High income: nonOECD	April 2013 database update: Based on data from the National Bureau o	
9	Argentina	ARG	Latin America & Caribbean	Upper middle income		
10	Armenia	ARM	Europe & Central Asia	Lower middle income		
11	American Samoa	ASM	East Asia & Pacific	Upper middle income		
12	Antigua and Barbuda	ATG	Latin America & Caribbean	High income: nonOECD	April 2012 database update: Based on official government statistics, na	
13	Australia	AUS	East Asia & Pacific	High income: OECD	Fiscal year end: June 30; reporting period for national accounts data: F	
14	Austria	AUT	Europe & Central Asia	High income: OECD	A simple multiplier is used to convert the national currencies of EMU n	
15	Azerbaijan	AZE	Europe & Central Asia	Upper middle income	April 2012 database update: National accounts historical expenditure s	
16	Burundi	BDI	Sub-Saharan Africa	Low income		
17	Belgium	BEL	Europe & Central Asia	High income: OECD	A simple multiplier is used to convert the national currencies of EMU n	
18	Benin	BEN	Sub-Saharan Africa	Low income		
19	Burkina Faso	BFA	Sub-Saharan Africa	Low income		
20	Bangladesh	BGD	South Asia	Low income		
21	Bulgaria	BGR	Europe & Central Asia	Upper middle income	Fiscal year end: June 30; reporting period for national accounts data: F	
22	Bahrain	BHR	Middle East & North Africa	High income: nonOECD	New reference year is 2010. April 2011 database update: The National S	
23	Bahamas, The	BHS	Latin America & Caribbean	High income: nonOECD	New base year is 2010.	
24	Bosnia and Herzegovina	BIH	Europe & Central Asia	Upper middle income		
25	Belarus	BLR	Europe & Central Asia	Upper middle income	New reference year is 2010.	

Figura 18. Pestanya "Metadata - Countries" de l'arxiu Excel sobre l'indicador de Població

	A	B	C	D	E	F	G	H	I	J	K
1	Data Source	World Development Indicators									
2											
3	Country Name	Country Code	Indicator Name	Indicator Code	1960	1961	1962	1963	1964	1965	1966
4	Aruba	ABW	Population, tota SP.POP.TOTL		54208	55435	56226	56697	57029	57360	57712
5	Andorra	AND	Population, tota SP.POP.TOTL		13414	14376	15376	16410	17470	18551	19646
6	Afghanistan	AFG	Population, tota SP.POP.TOTL		8774440	8953544	9141783	9339507	9547131	9765015	9990125
7	Angola	AGO	Population, tota SP.POP.TOTL		4965988	5056688	5150076	5245015	5339893	5433841	5526653
8	Albania	ALB	Population, tota SP.POP.TOTL		1608800	1659800	1711319	1762621	1814135	1864791	1914573
9	Arab World	ARB	Population, tota SP.POP.TOTL		93485943	96058179	98728995	101496308	104359772	107318159	110379639
10	United Arab Emirates	ARE	Population, tota SP.POP.TOTL		89608	97727	108774	121574	134411	146341	156890
11	Argentina	ARG	Population, tota SP.POP.TOTL		20623998	20959241	21295290	21630854	21963952	22293817	22618887
12	Armenia	ARM	Population, tota SP.POP.TOTL		1867396	1934239	2002170	2070427	2138133	2204650	2269475
13	American Samoa	ASM	Population, tota SP.POP.TOTL		20012	20478	21118	21883	22701	23518	24320
14	Antigua and Barbuda	ATG	Population, tota SP.POP.TOTL		54681	55403	56311	57368	58500	59653	60818
15	Australia	AUS	Population, tota SP.POP.TOTL		10276477	10483000	10742000	10950000	11167000	11388000	11651000
16	Austria	AUT	Population, tota SP.POP.TOTL		7047539	7086299	7129864	7175811	7223801	7270889	7322066
17	Azerbaijan	AZE	Population, tota SP.POP.TOTL		3897889	4030130	4167558	4307315	4445653	4579759	4708485
18	Burundi	BDI	Population, tota SP.POP.TOTL		2786740	2840375	2894510	2950903	3011957	3079034	3153879
19	Belgium	BEL	Population, tota SP.POP.TOTL		9153489	9183948	9220578	9289770	9378113	9463667	9527807
20	Benin	BEN	Population, tota SP.POP.TOTL		2431620	2466002	2503232	2543335	2586362	2632360	2681382
21	Burkina Faso	BFA	Population, tota SP.POP.TOTL		4829291	4894578	4960325	5027818	5098892	5174869	5256363
22	Bangladesh	BGD	Population, tota SP.POP.TOTL		49537147	50953503	52403243	53909233	55503132	57200412	59029826
23	Bulgaria	BGR	Population, tota SP.POP.TOTL		7867374	7943118	8012946	8078145	8144340	8204168	8258057
24	Bahrain	BHR	Population, tota SP.POP.TOTL		162501	167924	173107	178048	182774	187348	191782
25	Bahamas, The	BHS	Population, tota SP.POP.TOTL		109526	115108	121083	127331	133697	140049	146364

Figura 19. Pestanya "Data" de l'arxiu Excel sobre l'indicador de Població

6.2. Preparació de les dades

Treballar les dades des de 5 diferents fitxers és una tasca que resulta carregosa i feixuga, per tant s'ha considerat necessari crear un sol arxiu Excel amb tot el conjunt de dades ordenades.

Primer de tot, aclarir que les metadades adquirides del portal del Banc Mundial no només

incorporen tot el llistat de països, sinó que a més mesuren les estadístiques de certes regions, que no són pròpiament països. És el cas de les que veiem a continuació:

Country Name
Arab World
Central Europe and the Baltics
Caribbean small states
East Asia & Pacific
East Asia & Pacific (all income levels)
Europe & Central Asia
Europe & Central Asia (all income levels)
Euro area
European Union
Fragile and conflict affected situations
High income
Heavily indebted poor countries (HIPC)
Latin America & Caribbean
Latin America & Caribbean (all income levels)
Least developed countries: UN classification
Low income
Lower middle income
Low & middle income
Middle East & North Africa (all income levels)
Middle income
Middle East & North Africa
North America
High income: nonOECD
High income: OECD
OECD members
Other small states
Pacific island small states
South Asia
Sub-Saharan Africa
Sub-Saharan Africa (all income levels)
Small states
Upper middle income
World

Figura 20. Regions que s'han obviat dins dels arxius descarregats

relacionats amb l'ortografia dels noms.

Pel que fa a la tercera columna, correspon a la regió que se li assigna a cada país. El Banc Mundial adjudica cada país a una de 7 regions. Pel que fa al propòsit del projecte s'ha decidit prescindir d'algunes de les regions i deixar-ho en quatre, que després tindran un paper important en la interfície que es crearà. Així doncs, la distribució de regions que es farà serà la següent:

Com es pot veure en la figura, algunes de les regions que s'inclouen no són regions geogràfiques sinó econòmiques, és a dir engloba països amb nivells similars de riquesa. A part, mostra dades de zones geogràfiques com els països bàltics, l'Orient Mitjà, etc...

S'ha decidit obviar aquestes regions ja que per a la explotació i posterior representació que es vol crear, té més sentit centrar-se en els països del món. Així doncs, obviades aquestes regions, han quedat un total de 215 països dividits en diverses regions que es comentaran a posteriori.

La primera columna de l'arxiu, doncs, és el llistat de països. I la segona es tracta d'un codi per a cada país. Aquests codis no són aleatoris, sinó que provenen d'un estàndard estipulat per l'Organització Internacional de Normalització (ISO), concretament l'ISO 3166 que adjudica codis als noms dels països i a altres dependències administratives. Aquesta normalització té 3 variants: el codi numèric, l'alfa-2, i l'alfa-3. S'ha escollit l'alfa-3 (3 lletres). D'aquesta manera assurem que no manqui cap país, en facilitem la cerca, i s'eviten errors

Regions establertes	Regions en la pàgina del banc Mundial
Àfrica	Middle East & North Africa, Sub-Saharan Africa
Amèrica	Latin America & Caribbean, North America
Àsia	East Asia & Pacific, South Asia
Europa	Europe & Central Asia

Taula 5. Distribució de regions establerta

Degut a la manca de dades antigues en molts dels països, especialment els de poca població o en vies de desenvolupament, s'ha decidit escurçar el període d'estudi. S'ha decidit acotar el rang d'estudi entre l'any 1991 i el 2012, ja que 22 anys és una mesura suficientment característica de l'evolució dels indicadors.

Per tant, s'ha fet un muntatge tal que cada país estigués repetit tants cops com anys hi ha, és a dir 22. D'aquesta manera és possible incorporar la següent columna, els anys, i els 5 indicadors escollits: el Producte Interior Brut, la Mortalitat Infantil, l'Atur, l'Esperança de Vida, i la Població.

Com a incís, dir que finalment, per a la realització de la interfície s'ha decidit obviar l'indicador del PIB. S'ha considerat molt més característic i representatiu el PIB per càpita, que s'obté mitjançant la divisió de el PIB per la població. En cas d'haver escollit representar el PIB, aquest fet podria donar lloc a interpretacions errònies de les dades, ja que alguns països pobres tenen un PIB més elevat que altres degut a la major població que tenen. El PIB per càpita és qui dóna l'autèntica mesura del nivell econòmic d'un estat. Vegi's l'exemple següent:

PAÍS	PIB (\$)	POBLACIÓ	PIB p. CÀPITA (\$/pers.)
Bangladesh (1991)	30957444766,98	109934590	281,60
Andorra (1991)	1106890942,99	56674	19530,84

Taula 6. Comparativa entre el PIB i el PIB per càpita entre dos països seleccionats

Si representéssim les dades de PIB no s'obtidrien resultats coherents, com es pot apreciar en aquesta taula. A l'any 1991 Bangladesh tenia un PIB molt més elevat que Andorra perquè aquest últim és un dels països amb menys població de tot el món. En canvi, si s'analitza el PIB per càpita, queda palesa la situació econòmica de Bangladesh, un país que ja el 1991 estava sumit en la pobresa i que avui dia encara ho està.

Així doncs, l'arxiu de dades creat incorporarà la columna del PIB per càpita.

La idea de tot plegat és la de poder facilitar la cerca de dades concretes mitjançant filtres establerts a cada columna. El resultat és el següent:

	Country Name	Country Code	Region	Year	PIB	MORTALITAT INFANT	ATUR	ESPERANÇA VIDA	POBLACIÓ	PIBpercapita
1	Aruba	ABW	America	1991				73,43312195	64623	
2	Afghanistan	AFG	Asia	1991		171,5	8,69999998	49,40456098	12612043	
3	Angola	AGO	Africa	1991	12193746624,678879	225,8	7,59999999	41,20634146	10652727	1144,659637
4	Albania	ALB	Europa	1991	1139166676,0109425	38,8	14,3	71,91841463	3266790	348,711327
5	Andorra	AND	Europa	1991	1106890942,9884675	7,9			56674	19530,84206
6	United Arab Emirates	ARE	Africa	1991	51552165568,876045	15,7	2,5	72,0137561	1908002	27018,92638
7	Argentina	ARG	America	1991	189719989668,1026	26,6	5,8000002	71,76329268	33075194	5736,020465
8	Armenia	ARM	Europa	1991	2068526521,9029896	47,4	20	67,73829268	3511912	589,0029482
9	American Samoa	ASM	Asia	1991					48379	
10	Antigua and Barbuda	ATG	America	1991	410388881,64115983	24,2		71,40058537	62412	6575,480383
11	Australia	AUS	Asia	1991	325641629385,44855	8,6	9,6000004	77,27560976	17284000	18840,64044
12	Austria	AUT	Europa	1991	173375508073,07013	9	3,4000001	75,56780488	7754891	22356,92392
13	Azerbaijan	AZE	Europa	1991	8792365810,509403	94,8	6	64,42414634	7271000	1209,237493
14	Burundi	BDI	Africa	1991	1167398478,3459036	171,9	7,69999998	46,9014878	5749774	203,0338024
15	Belgium	BEL	Europa	1991	211312782752,5103	9,6	7	76,19219512	10004486	21121,80303
16	Benin	BEN	Africa	1991	1986437797,8767164	174,3	0,8	53,93570732	5182525	383,2953624
17	Burkina Faso	BFA	Africa	1991	3135045586,3077	201,8	2,59999999	49,31646341	9050090	346,410432
18	Bangladesh	BGD	Asia	1991	30957444766,984383	137,6	3,59999999	60,53182927	109934590	281,5987649
19	Bulgaria	BGR	Europa	1991	10943548723,760754	22,6	18,5	71,56097561	8632367	1267,734414
20	Bahrain	BHR	Africa	1991	4616223477,212885	22,6	6,3000002	72,7095122	509645	9057,723469
21	Bahamas, The	BHS	America	1991	3111160000	22,5	12,2	70,79180488	261102	11915,49663
22	Bosnia and Herzegovina	BIH	Europa	1991		17	22,2999999	66,37626829	4373715	
23	Belarus	BLR	Europa	1991	17813389815,003983	16,6	10,1	70,37804878	10194000	1747,438671
24	Belize	BLZ	America	1991	444720726,46485	37,2	10,9	70,98195122	191368	2323,903299
25	Bermuda	BMU	America	1991	1634899968			74,0295122	59021	27700,30952
26	Bolivia	BOL	America	1991	5343259388,763367	118	4,3000002	59,30241463	6956736	768,0698806
27	Brazil	BRA	America	1991	407337832905,40546	58,9	10	66,90997561	152153539	2677,149908

Figura 21. Arxiu Excel definitiu amb una disposició òptima per a treballar les dades

7. Creació de la interfície

Un cop s'han descarregat les dades i s'han treballat de manera que estiguin en un mateix Excel i organitzades tal i com s'ha explicat en el punt anterior, cal explotar-les fent servir R i crear la interfície visual desitjada amb Shiny. Per això s'ha considerat necessari fer una descripció sobre el funcionament de l'eina Shiny.

7.1. Introducció a Shiny

Shiny és un paquet que serveix per construir aplicacions web ("apps") interactives directament des del programa R. Per poder-ne fer ús cal disposar de connexió a internet i instal·lar el paquet mitjançant la comanda següent en la pantalla de comandes de R:

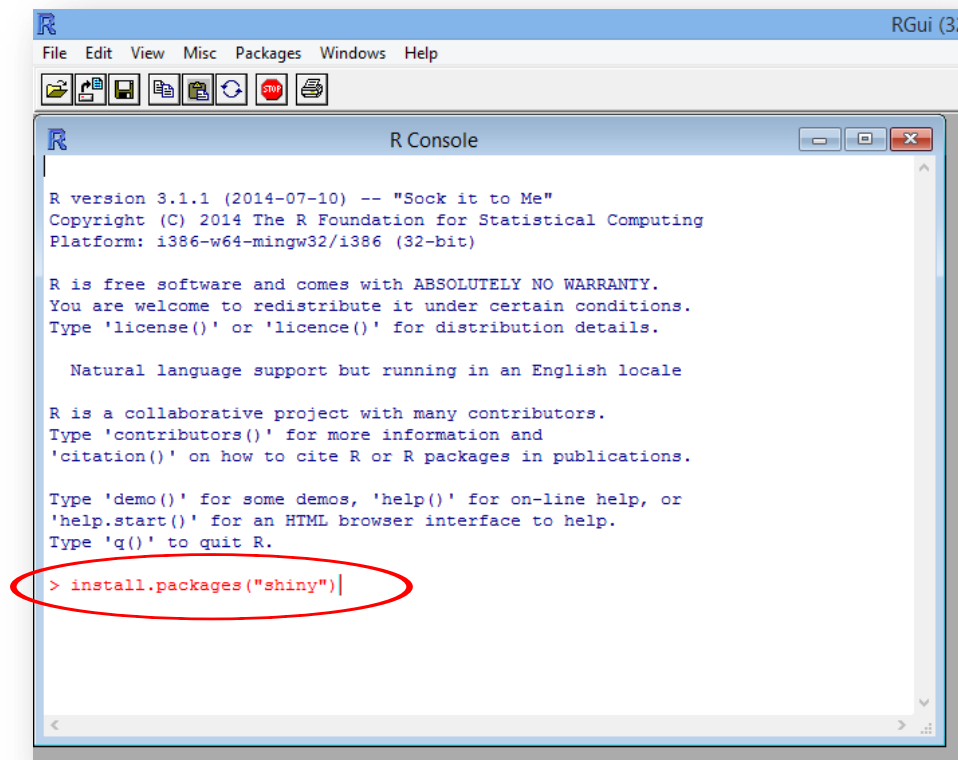


Figura 22. Consola de programació de R. Comanda per instal·lar el paquet Shiny.

El funcionament bàsic d'una aplicació de Shiny és relativament senzill, ja que només consisteix en dos components:

- Un arxiu anomenat **ui.R**, que és qui controla l'aparença i l'estructura de l'aplicació.

- Un arxiu anomenat **server.R**, que conté la informació que l'ordinador necessitarà per construir l'aplicació.

Veiem un exemple senzill d'una aplicació creada amb Shiny a partir de dades sobre la duració de les erupcions de Guèisers:

El seu arxiu **ui.R** determina que l'usuari que faci ús de l'aplicació i en vegi el seu gràfic podrà escollir (canviar) el número de barres que veu en el gràfic (10, 20, 35, o 50), podrà veure les observacions individuals o no, i a més a més podrà canviar característiques de l'eix Y si tria veure els valors estimats de densitat.



```

server.R  ui.R

shinyUI(bootstrapPage(

  selectInput(inputId = "n_breaks",
    label = "Number of bins in histogram (approximate):",
    choices = c(10, 20, 35, 50),
    selected = 20),

  checkboxInput(inputId = "individual_obs",
    label = strong("Show individual observations"),
    value = FALSE),

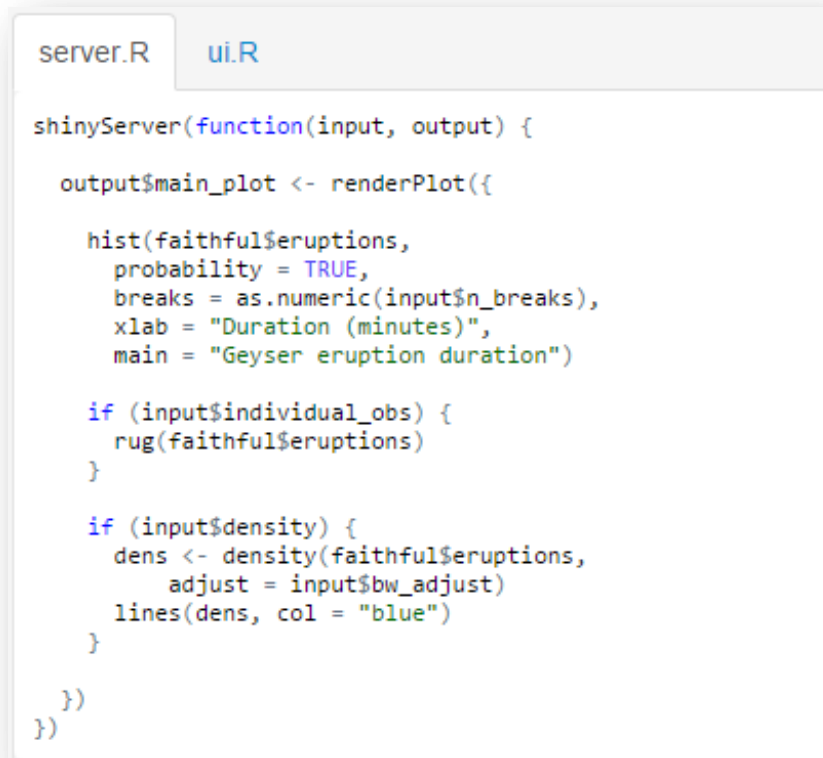
  checkboxInput(inputId = "density",
    label = strong("Show density estimate"),
    value = FALSE),

  plotOutput(outputId = "main_plot", height = "300px"),

  # Display this only if the density is shown
  conditionalPanel(condition = "input.density == true",
    sliderInput(inputId = "bw_adjust",
      label = "Bandwidth adjustment:",
      min = 0.2, max = 2, value = 1, step = 0.2)
  )
))
  
```

Figura 23. Exemple de fitxer de R Ui.R per a una app de Shiny. Font: <http://shiny.rstudio.com/>

El **server.R** és l'encarregat de fer possible que l'usuari interactuï amb l'aplicació. És a dir, dibuixa el gràfic desitjat relacionant-lo amb les variables definides al **ui.R**, de manera que cada cop que l'usuari canviï una opció (per exemple el número de barres que es veuen en el gràfic) això tindrà efecte immediat a l'aplicació i es canviarà el nombre de barres.



```
server.R      ui.R

shinyServer(function(input, output) {

  output$main_plot <- renderPlot({

    hist(faithful$eruptions,
         probability = TRUE,
         breaks = as.numeric(input$n_breaks),
         xlab = "Duration (minutes)",
         main = "Geyser eruption duration")

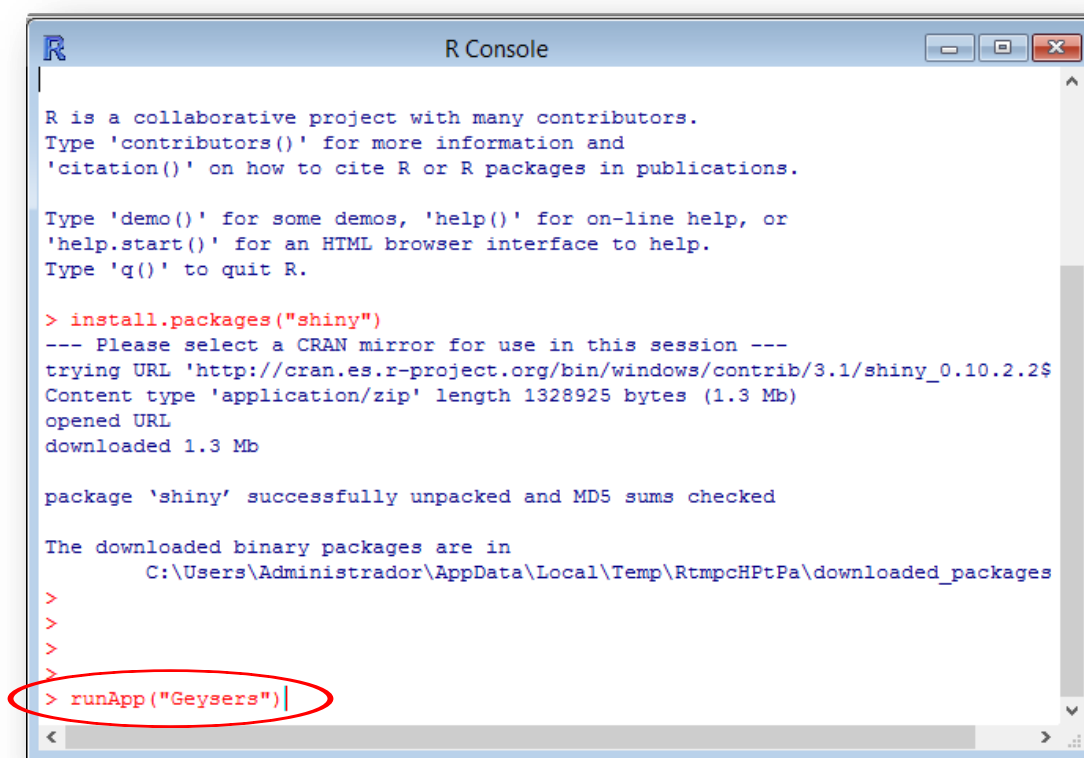
    if (input$individual_obs) {
      rug(faithful$eruptions)
    }

    if (input$density) {
      dens <- density(faithful$eruptions,
                      adjust = input$bw_adjust)
      lines(dens, col = "blue")
    }

  })
})
```

Figura 24. Exemple de fitxer Server.R per a una app de Shiny. Font: <http://shiny.rstudio.com/>

Així doncs, caldrà que guardem aquests dos fitxers en una carpeta amb el nom desitjat per a la nostra aplicació, per exemple "Geysers". Aquesta carpeta haurà d'estar al directori on estiguem treballant amb R. Per posar en marxa l'aplicació només haurem de cridar la següent comanda en la consola de R:



```
R
R Console

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> install.packages("shiny")
--- Please select a CRAN mirror for use in this session ---
trying URL 'http://cran.es.r-project.org/bin/windows/contrib/3.1/shiny_0.10.2.2$
Content type 'application/zip' length 1328925 bytes (1.3 Mb)
opened URL
downloaded 1.3 Mb

package 'shiny' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
  C:\Users\Administrador\AppData\Local\Temp\RtmpcHPTPa\downloaded_packages

>
>
>
>
> runApp("Geysers")
```

Figura 25. Consola de programació de R. Comanda per posar en marxa l'aplicació creada amb Shiny

A continuació s'obrirà l'aplicació en el navegador que tinguem configurat per defecte. El resultat serà el següent:

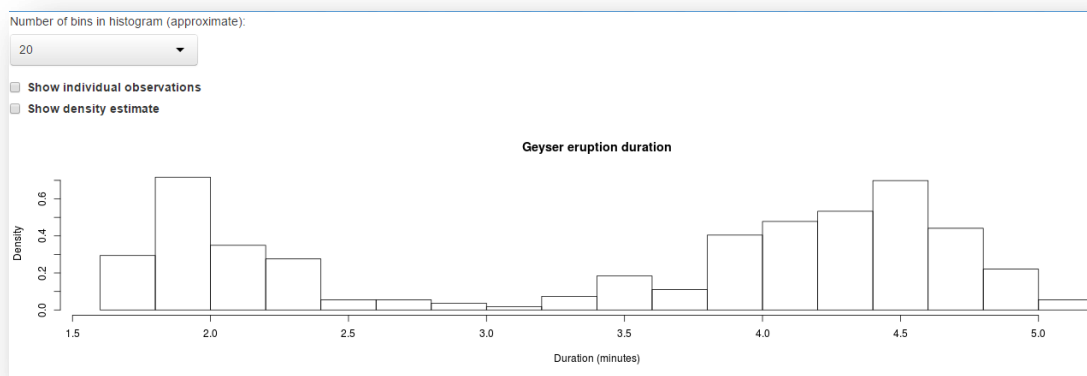


Figura 26. Exemple de gràfic de Shiny. Font: <http://shiny.rstudio.com/>

7.2. Desenvolupament i creació de la interfície

La interfície final pretén ser un explorador socioeconòmic de la situació mundial en els darrers 22 anys.

Els indicadors seleccionats, que ja s'han comentat al llarg dels punts anteriors, s'han considerat suficientment representatius per a que un usuari que faci ús de la interfície pugui extreure conclusions acurades de no només la situació socioeconòmica actual sinó de l'evolució d'aquesta al llarg dels darrers anys.

La interfície està creada amb el paquet Shiny del programa R. El programari R permet fer ús de l'aplicació tant dins del propi programa com en el navegador web establert per defecte. Tanmateix, l'usuari no necessitarà tenir instal·lat el programa, ni requerirà tenir coneixements previs de programació, ni tan sols d'estadística. Es tracta que visualment, des del seu navegador web, pugui visitar la interfície i mitjançant les opcions que se li presenten pugui adquirir suficient informació com per arribar a certes conclusions.

L'explorador socioeconòmic constarà de 3 pestanyes, cada una amb una funció diferent, que l'usuari podrà explorar tantes vegades com necessiti i podrà interactuar amb elles de la forma en la que s'explicarà en els següents apartats. Les 3 pestanyes seran aquestes:

- Explorador de dades
- Anàlisi de relacions
- Anàlisi geogràfic

Abans d'explicar en que consistirà cada una de les pestanyes és important fer un petit incís per definir el mot anglès "widget", ja que és un mot que s'usa amb freqüència en els següents apartats:

Un "**widget**" és una petita aplicació o programa, sovint dins d'una aplicació més gran, que té la tasca de facilitar l'accés a funcions que són usades amb freqüència dins l'aplicació i que sol aportar informació visual.

7.2.1. Explorador de dades

La primera pestanya de la interfície és la menys espectacular a nivell visual, però no per això menys important.

L'explorador de dades té format de taula i té la tasca de facilitar la cerca i la visualització de dades concretes a l'usuari. De fet, aquesta primera pestanya correspon a l'arxiu Excel que s'ha creat prèviament, però presentat d'una forma més interactiva. Els elements que conté són els següents:

- **La taula:** la taula agrupa totes les dades seleccionades, de manera que tingui les columnes País, Any, PIB per càpita, mortalitat infantil, atur, esperança de vida, i

població. Cada una d'aquestes columnes es podrà ordenar (alfabèticament la del llistat de països, cronològicament la dels anys, i segons els seus valors les dels indicadors).

- **Dos "widgets" de control a mode de desplegable:** aquests dos "widgets" consisteixen en dos requadres que contenen diferents opcions que l'usuari podrà triar. El primer és per als països, i permet seleccionar un país, de manera que immediatament a la taula no aparegui cap informació relativa als països que no són aquell. El segon widget és per la data, de manera que l'any que s'esculli farà que es restringeixi la informació de la taula a aquell any.
- **Un "widget" de cerca que rep un input de text:** aquest requadre de cerca va acompanyat de la paraula "Search" i permet escriure un nom o un codi ISO de país, per restringir la cerca.
- **Un "widget" de control a mode de desplegable:** aquest simplement serveix per triar el nombre de columnes que un desitja que es mostren per pàgina. 10, 25, 50 o 100

Figura 27. Widgets incorporats en l'explorador de dades

D'aquesta manera si l'usuari vol cercar, per exemple, els valors de tots els indicadors relatius a Espanya, l'any 2004 ho pot fer senzillament, mitjançant els dos seleccionables o escrivint:

country.name	country.code	region	year	mortalitat.infantil	atur	esp.vida	poblacio	PIBpercapita
Spain	ESP	Europa	2004	5.9	11.1	79.87073	42921895	24919.73

Figura 28. Dades relatives a Espanya, l'any 2004

La pestanya de l'explorador de dades, doncs, té la aparença final següent:

country.name	country.code	region	year	mortalitat.infantil	atur	esp.vida	poblacio	PIBpercapita
Aruba	ABW	America	2012			75.20676	102384	
Afghanistan	AFG	Asia	2012	100.0	8.6	60.50912	29824536	687.5814
Angola	AGO	Africa	2012	172.6	7.5	51.46400	20820525	5539.8007
Albania	ALB	Europa	2012	15.4	14.7	77.35046	2801681	4406.1153
Andorra	AND	Europa	2012	3.1			78360	
United Arab Emirates	ARE	Africa	2012	8.5	3.8	76.95788	9205651	40444.0670
Argentina	ARG	America	2012	13.7	7.2	76.01268	41086927	14679.9252
Armenia	ARM	Europa	2012	16.4	18.5	74.43722	2969081	3353.9731
American Samoa	ASM	Asia	2012				55128	
Antigua and Barbuda	ATG	America	2012	9.6		75.66532	89069	13525.6162

Figura 29. Aparença final pestanya de l'explorador de dades

7.2.2. Anàlisi de relacions

La pestanya d'anàlisi de relacions actua com a gràfic interactiu, que pretén estudiar quines relacions existeixen entre les variables. Mitjançant un gràfic de dos eixos (X i Y), l'usuari podrà escollir dos dels indicadors i descobrir si estan relacionats entre ells, a més a més de l'evolució cronològica respectiva. Elements dels quals es compon aquesta pestanya:

- **Un diagrama bivariant:** on a l'eix Y hi apareix un indicador i en l'eix X un altre. El gràfic estudia la interrelació entre els dos indicadors seleccionats, per un any concret. Hi apareix un punt per a cada país. Com més gran sigui la població del país més gran es veu el punt en el gràfic. Els colors dels punts varien depenent de les regions geogràfiques en les quals s'han distribuït els països (veure punt 6.2) segons la següent llegenda:

Regions establertes	Color en el gràfic bivariant
Àfrica	vermell
Amèrica	blau
Àsia	verd
Europa	lila

Taula 7. Llegenda per al gràfic bivariant

A l'esquerra del gràfic hi haurà un panell amb els següents "widgets":

- **Dos "widgets" de control a mode de desplegable:** es tracta de dos requadres que permeten a l'usuari escollir quin indicador vol veure representat en l'eix X, i quin en l'eix Y. És a dir es podrà triar entre: PIB, mortalitat, atur, i esperança de vida.
- **Un "widget" de control amb etiquetes:** aquest requadre és similar a un desplegable, però també pot rebre un input de text. El que fa és que permet a l'usuari seleccionar un o més països de manera que el nom del país seleccionat apareixerà sobre el punt corresponent en el diagrama bivariant. Vegi's l'exemple següent corresponent a Mònaco, on apareix la mortalitat infantil a l'eix Y i el PIB per càpita a l'eix X:

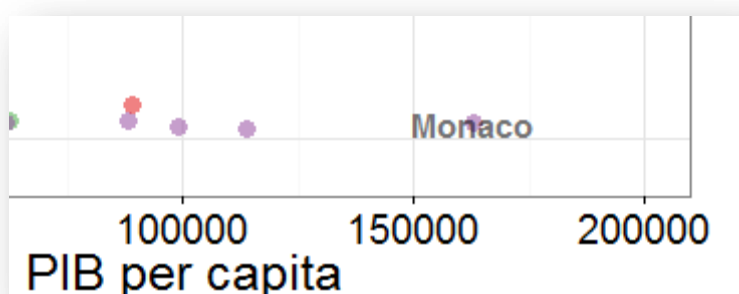


Figura 30. Funcionament del "widget" amb etiquetes

- **Un "widget" de control relacionat amb la cronologia:** aquest "widget" consisteix en un interval que va des de l'any 1991 fins al 2012. L'usuari podrà clicar en un any i el gràfic mostrarà els valors dels indicadors interrelacionats en aquell any. A més a més, incorpora un botó "play", que si es clica mostra l'evolució de la relació entre indicadors des de l'any que s'hagi seleccionat fins al final.

El panell amb els "widgets" te l'aparença següent:

Variable X:
PIB per capita

Variable Y:
mortalitat infantil

Mostrar etiquetes dels següents països:
Burkina Faso

Any:
1,991 2,011

▶

Figura 31. Panell de "widgets" que acompanya el gràfic bivariant

Per altra banda, l'aspecte visual del gràfic bivariant corresponent a les opcions seleccionades en el panell anterior és tal i com el que es mostra a la figura següent:

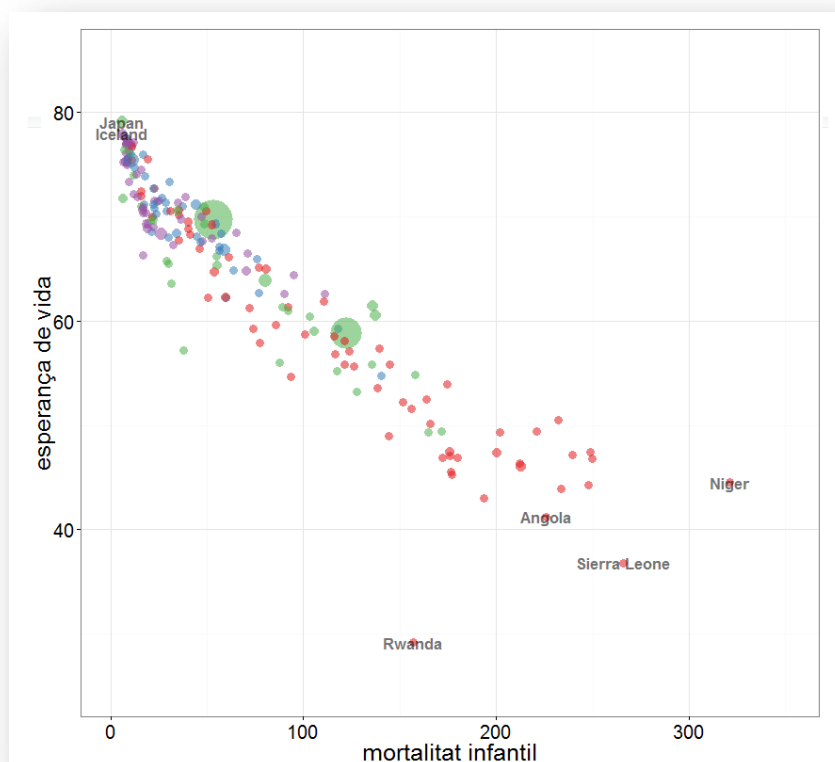


Figura 32. Gràfic bivariant

7.2.3. Anàlisi geogràfic

La pestanya d'anàlisi geogràfic, com el seu nom indica, consisteix en un mapa que pretén dotar a l'usuari d'una percepció visual de l'indicador en un cert any.

Els elements dels quals es compon són els següents:

- **Un "widget" de control a mode de desplegable:** es tracta d'un requadre que permet a l'usuari escollir quin dels indicadors vol veure representat en el mapa. És a dir es podrà triar entre: PIB, mortalitat, atur, i esperança de vida. A cada un dels indicadors se li ha atorgat un color diferent mitjançant el codi de programació.

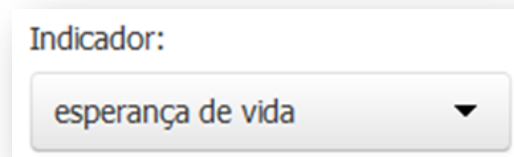


Figura 33. Widget de control desplegable

- **Un "widget" de control relacionat amb la cronologia:** aquesta és una de les parts més interessants d'aquesta pestanya. Aquest "widget" consisteix en un interval que va des de l'any 1991 fins al 2012. L'usuari podrà clicar en un any i el mapa mostrarà els valors de l'indicador en aquell any. A més a més, incorpora un botó "play", que si es clica mostra l'evolució de l'indicador des de l'any que s'hagi seleccionat fins al final.



Figura 34. Widget de control cronològic

- **Un mapa interactiu:** la part fonamental d'aquesta pestanya és el mapa. Ocupa gairebé tota la pantalla. Els seus colors aniran canviant a mesura que es canviï l'indicador seleccionat i les tonalitats dels colors variaran segons els valors de l'indicador estudiat.
- **Una llegenda:** la llegenda apareixerà a sota del mapa i mostra, per a cada indicador i per a cada any, el rang de valors de l'indicador (des del mínim d'aquell any fins al màxim) i les diferents tonalitats que adoptarà el mapa. Com més alt el valor de

l'indicador, més fosca la tonalitat del color; i com més baix, més propera al color blanc.

- **Un "widget" de control a mode de casella seleccionable:** aquesta casella resulta molt útil, ja que en clicar-la el que fa és fixar el rang de la llegenda. És a dir, estableix un mínim i un màxim per a tot el conjunt d'anys, d'aquesta manera quan l'usuari clica el "play", es pot veure l'evolució cronològica de l'indicador per tot el mapa. Si no es selecciona, el rang varia per cada any, sent el valor mínim el menor en aquell any, i el màxim el major en aquell any. Així doncs, si l'usuari "clica" el play percebrà variacions il·lògiques, que no entendrà, degudes al fet que els colors varien ja que el rang per a cada any, i per tant els intervals de tonalitats, no són els mateixos.

☒ Colorejar en termes absoluts?

Figura 35. Widget seleccionable

A continuació es mostra l'aparença d'aquesta pestanya per a cada indicador, havent seleccionat l'any 2012.

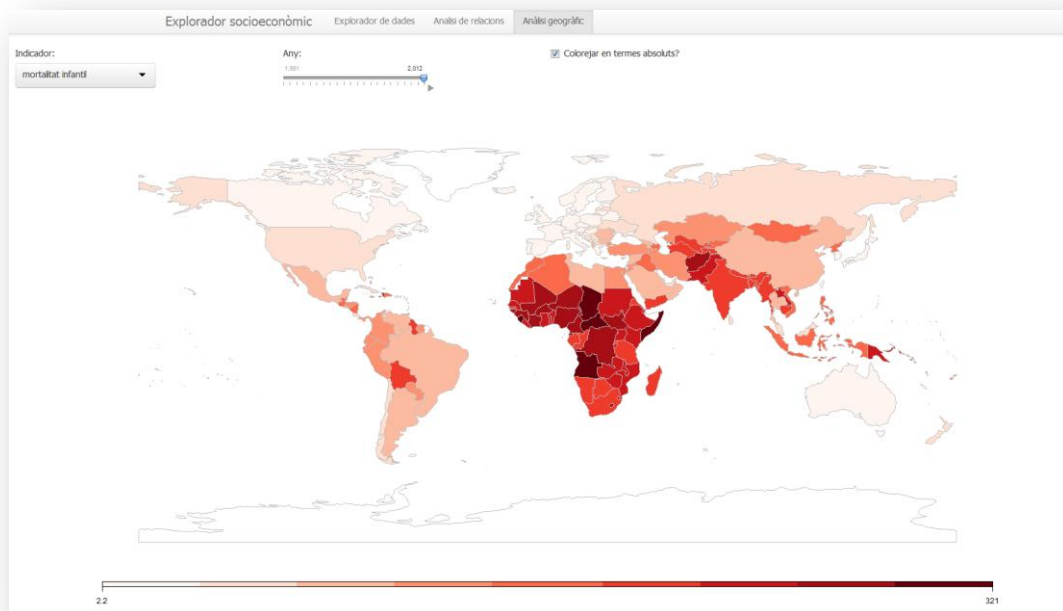


Figura 36. Mapa visual de l'indicador "mortalitat infantil" l'any 2012

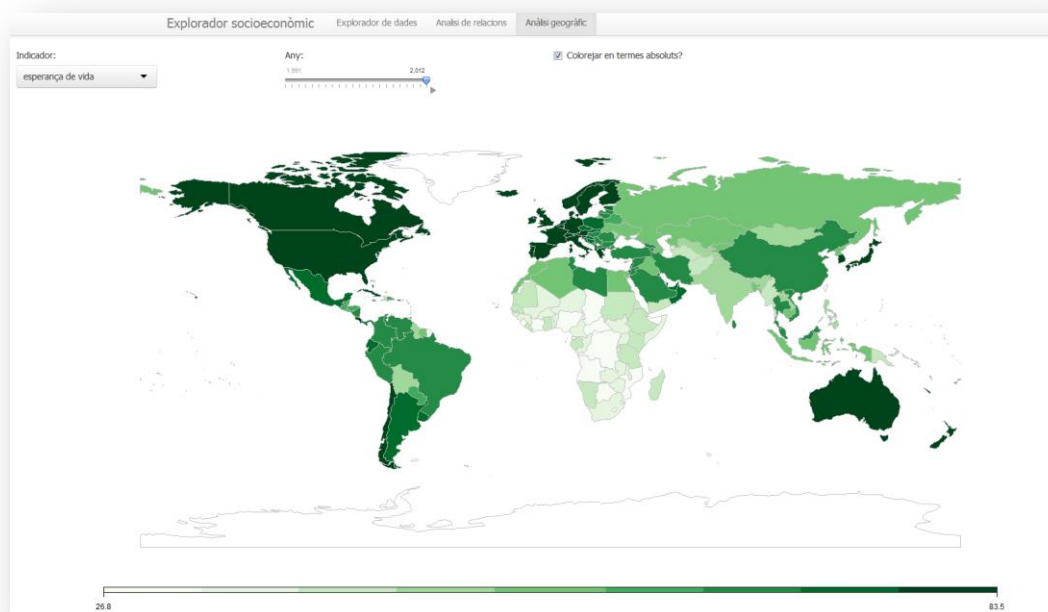


Figura 37. Mapa visual de l'indicador "esperança de vida" l'any 2012

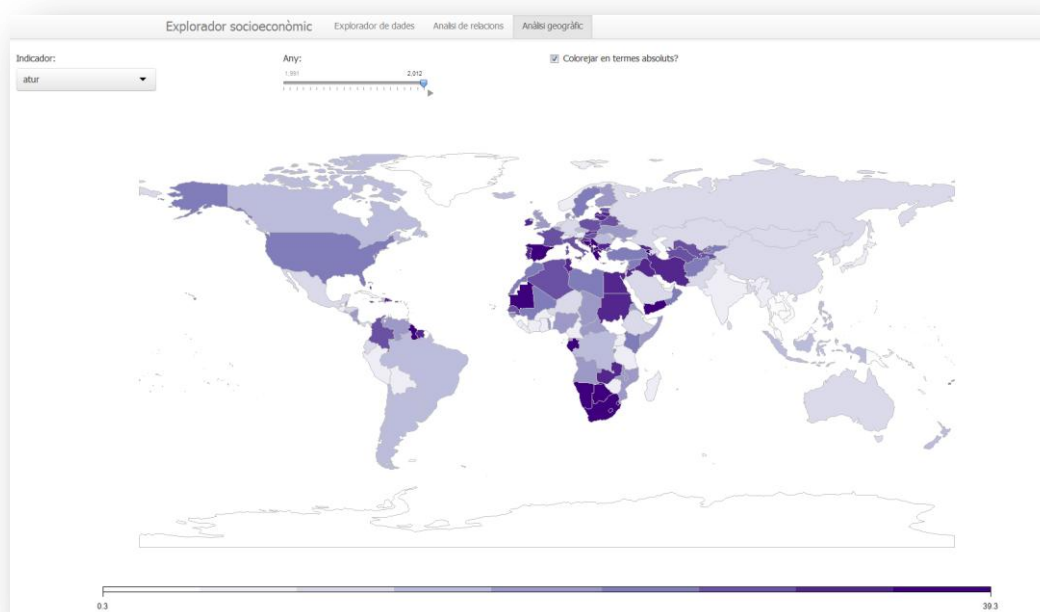


Figura 38. Mapa visual de l'indicador "atur" l'any 2012

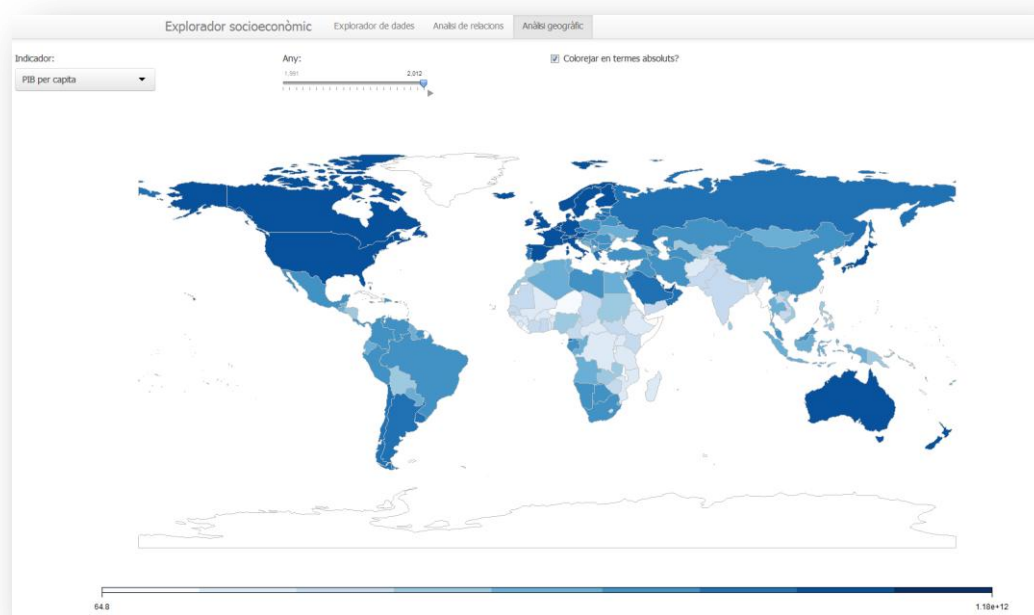


Figura 39. Mapa visual de l'indicador "PIB per càpita" l'any 2012

8. Explotació de la interfície

Un dels objectius en crear la interfície era que les dades que mostrés fossin representatives de l'evolució dels indicadors i que permetessin a un usuari ocasional, no expert en estadística, extreure conclusions concises sobre els conjunts de dades. És per aquest motiu que s'ha considerat necessari extreure alguna conclusió clara sobre les representacions de la interfície. No es pretén fer un anàlisi socioeconòmic detallat, simplement es volen demostrar els dos punts següents:

- La interfície funciona correctament. Les representacions no són aleatòries sinó que corresponen a les dades extretes dels fitxers. Els resultats que s'obtenen són lògics (comparats amb estàndards de riquesa i pobresa que hom coneix) i per tant l'anàlisi que se'n pot fer és vàlid.
- No es requereixen coneixements en estadística per fer una interpretació acurada de les dades. Les pestanyes creades presenten d'una manera visual les dades permetent a l'usuari veure l'evolució dels indicadors al llarg del temps i les relacions que tenen entre ells.

Així doncs, es farà esment d'alguna de les conclusions bàsiques que es pot treure a partir de la interacció amb l'explorador socioeconòmic creat.

L'anàlisi geogràfic permet observar amb només un cop de vista l'evolució d'algun dels indicadors. És útil, en aquest cas, seleccionar la casella de "colorejar en termes absoluts". D'aquesta manera es fixa el rang de valors per a tots els anys. Per exemple, si un selecciona l'indicador "Mortalitat Infantil", i fa una captura de pantalla per l'any 1991 i una per l'any 2012 pot extreure les següents conclusions:

Al 1991 el continent africà registra, amb diferència, els valors més negatius pel que fa a mortalitat infantil, amb països que arriben als 300 morts (de 5 anys o menys) per a cada 1000 naixements. Europa és el continent que presenta els registres més positius. Malauradament, al 2012 Àfrica segueix sent el continent amb els registres més elevats de mortalitat infantil. La nota positiva és que s'observa una tendència a la baixa, no només a Àfrica sinó a nivell mundial. Però els països en vies de desenvolupament encara registren uns valors molt elevats, en comparació als països occidentals, amb xifres que superen els centenars de víctimes infantils.

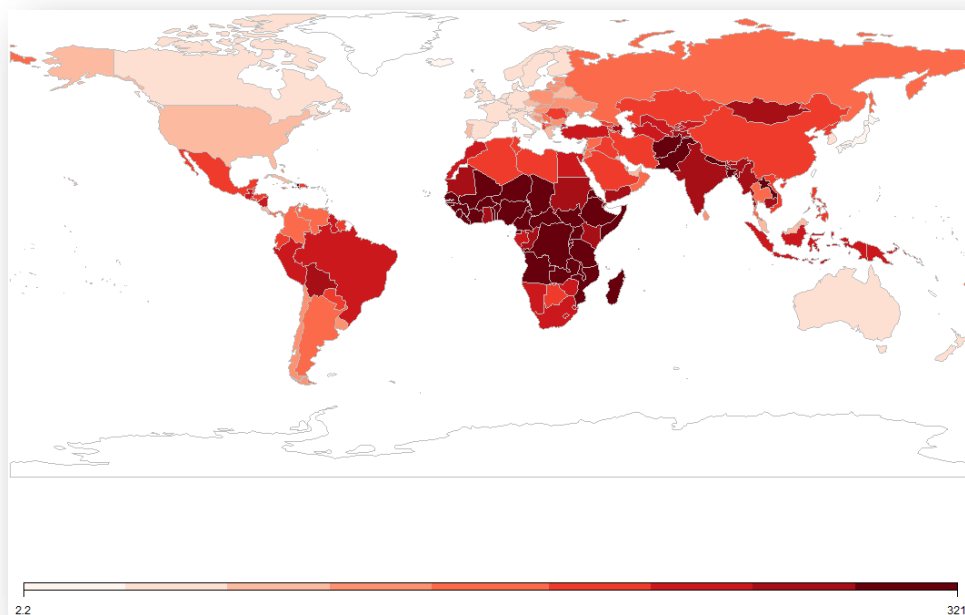


Figura 40. Mortalitat infantil al món, l'any 1991

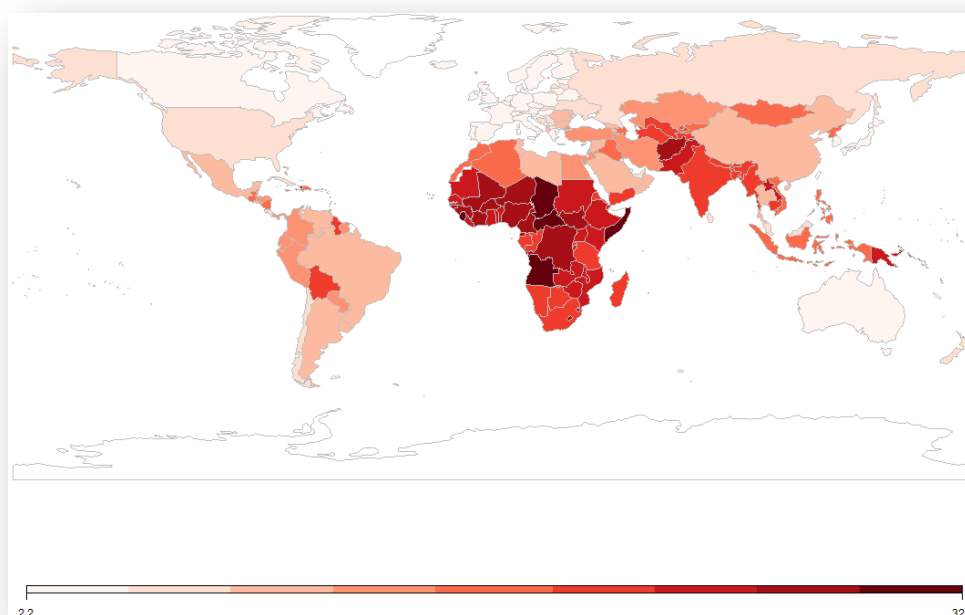


Figura 40. Mortalitat infantil al món, l'any 2012

Pel que fa al diagrama bivariant, també es poden treure conclusions clares. En aquest cas s'ha seleccionat l'esperança de vida per a l'eix Y, i la mortalitat infantil per a l'eix X. S'han tornat a fer dues captures de pantalla, una relativa a l'any 1991 i una al 2012 per veure quina és la tendència.

Primer de tot, i d'acord amb el que s'ha vist al mapa, s'aprecia que la majoria de països

africans (representats en vermell) són els que presenten la mortalitat infantil més elevada. A més, registren l'esperança de vida menor. El cas oposat que els països occidentals. Al 2012, els pitjors valors segueixen sent pertanyents a punts vermells. S'han etiquetat diversos països, per poder estudiar la seva evolució (s'han etiquetat els que registraven els pitjors i millors valors per a cada un dels indicadors i per a les dues dates), Hong Kong, Islàndia, Japó, Ruanda, Níger, Sierra Leone i Angola.

Un cop més, la nota positiva es que la tendència és a l'alça pel que fa a l'esperança de vida i a la baixa pel que fa a la mortalitat infantil. Els punts es desplacen cap a l'esquerra i cap amunt dins del diagrama bivariant.

Si un es fixa en els països etiquetats, els que estaven amunt ho segueixen estant i encara més (és el cas de Hong Kong, Islàndia i Japó), i pel que fa als africans seleccionats el que presenta la millora més significativa quant a l'esperança de vida és Ruanda i quant a la mortalitat infantil el Níger.

Destacar que aquest diagrama deixa palès (com ja se sabia) que l'esperança de vida i la mortalitat infantil són dos conceptes que estan interrelacionats i que com més elevat és un, més baix sol ser l'altre.

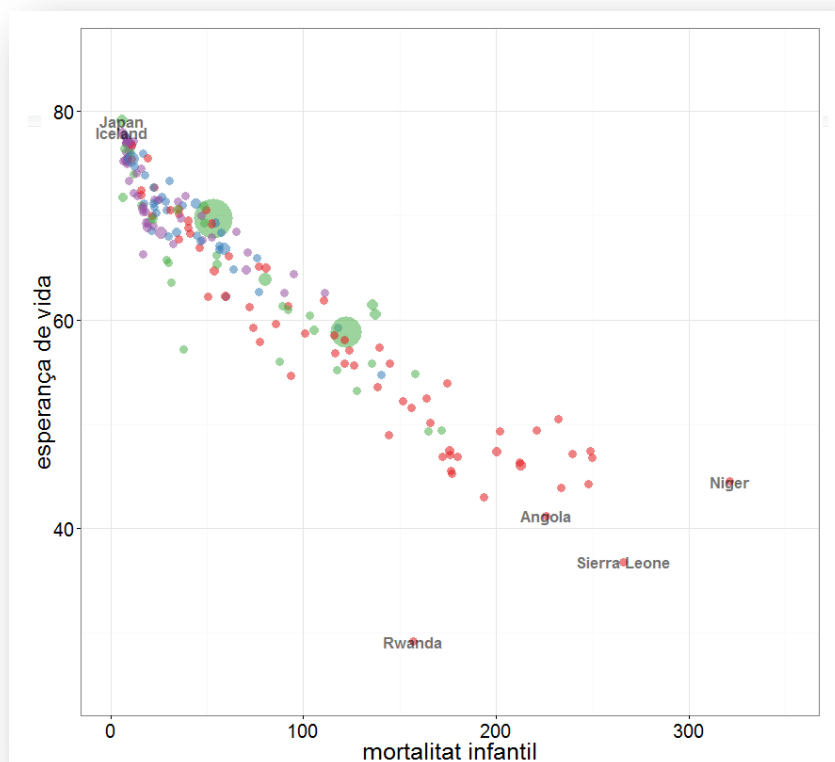


Figura 41. Diagrama bivariant. Any 1991.

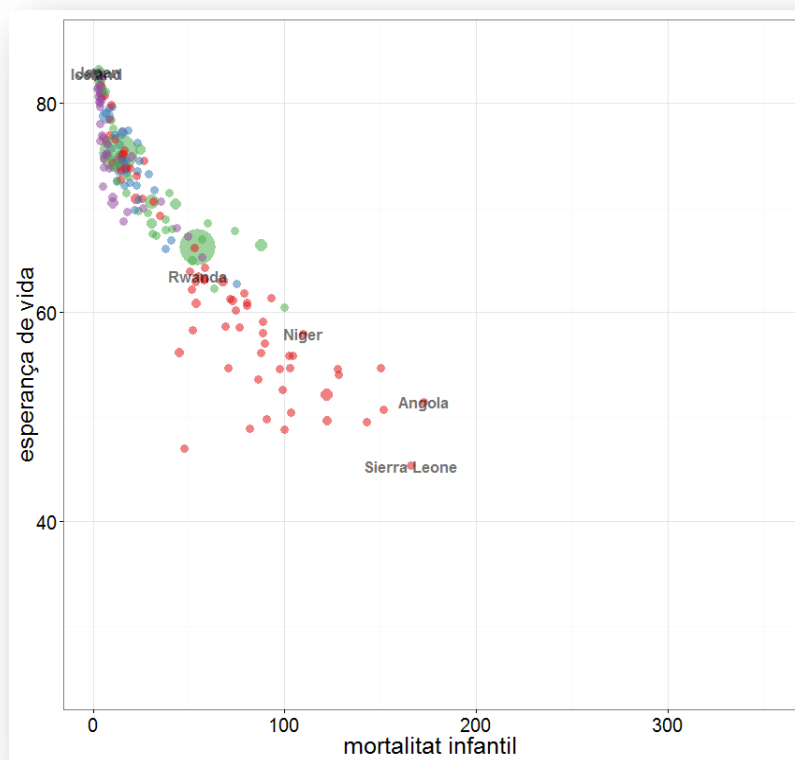


Figura 42. Diagrama bivariant. Any 2012.

Amb aquesta petita mostra del potencial de la interfície es conclou, doncs, que les conclusions que s'extreuen a partir de la interacció amb l'aplicació són acurades. Els exemples seleccionats han estat exemples senzills, dels quals en poguéssim pronosticar els resultats amb facilitat. Ha representat les dades amb èxit, per tant, l'aplicació funciona, i se n'ha demostrat la utilitat. A partir d'aquí és tasca de l'usuari explotar-la com desitgi.

9. Conclusions

9.1. Revisió dels objectius plantejats

Un cop realitzat el projecte, els objectius plantejats a l'inici del mateix es consideren satisfactòriament assolits.

En primer lloc, la contextualització sobre l'Open Data ha estat bàsica per a instruir-se sobre el moviment, i permet a un lector sense coneixements en el tema entendre quins beneficis aporta no només a les administracions sinó també a la ciutadania.

Pel que fa a la recerca d'iniciatives Open Data, els resultats han estat àmpliament satisfactoris ja que s'ha descobert que arreu del món hi ha centenars de portals amb catàlegs de dades en diferents formats. En el treball s'han adjuntat "links" d'alguns dels més destacats. Ha estat una grata sorpresa descobrir que a l'àmbit nacional cada cop hi ha més ajuntaments i entitats que estan obrint portals que contenen dades de rellevància, cosa que indica que aquest moviment no està estancat, ans el contrari, està en fase de creixement i la quantitat de projectes que s'estan duent a terme a partir de l'explotació de dades arriba a xifres considerables.

La recerca de portals que presentin les dades en forma d'interfícies visuals era el punt que es presentava amb més dubte, ja que es desconeixia l'abast i la quantitat dels portals existents. Finalment però, no només no hi ha hagut problema per trobar-ne, sinó que els que s'han trobat han resultat ser de gran utilitat i una font d'inspiració i ajuda per a la posterior creació d'una interfície pròpia.

Quant a la creació de la interfície visual, s'han seleccionat diverses opcions que semblaven factibles a l'hora d'elaborar-la i posteriorment a l'estudi corresponent la més favorable ha estat el paquet Shiny de RStudio.

A partir de l'aprenentatge del llenguatge de programació R i del seu paquet Shiny s'ha aconseguit crear una interfície el resultat de la qual es considera un èxit. Tenint en compte que el nivell de coneixement de programació amb R abans d'iniciar el projecte era nul, la interfície obtinguda és visualment bastant espectacular. Es tracta d'un explorador socioeconòmic que permet a un usuari fer un anàlisi geogràfic i un anàlisi de relacions a partir de 5 indicadors socioeconòmics de rellevància. El funcionament de la interfície està explicat en detall en el treball.

El fet d'haver creat una interfície visual amb èxit aporta una perspectiva molt positiva pel que fa als propòsits que es poden assolir mitjançant l'explotació de dades obertes i deixa la porta

oberta a crear altres tipus d'aplicacions en un futur proper, que poden tenir sentit tant en l'àmbit educatiu, com en l'àmbit professional, i fins i tot a nivell de hobby.

9.2. Línies d'evolució futures de l'aplicació

Malgrat el resultat satisfactori obtingut de l'aplicació, és cert que hi ha aspectes que es podrien millorar per tal d'incrementar la interacció entre aplicació i usuari i fer la interfície encara més atractiva.

Principalment, els aspectes a millorar tenen a veure amb el codi de programació, i estic segur que amb temps i dedicació es podrien aconseguir.

Per exemple, en l'anàlisi geogràfic seria molt interessant que a l'usuari, en passar el ratolí per sobre d'un país, l'hi aparegués un requadre amb informació referent a aquell país. Aquest requadre podria contenir el nom del país i el valor, en xifres, de l'indicador escollit per aquell any. Això facilitaria la tasca de comprensió a l'usuari, ja que és possible que no conegui la localització de tots els països en el mapa mundi, a més a més que potser li interessa conèixer els valors exactes en el mapa, en comptes dels intervals.

Pel que fa a l'anàlisi de relacions, l'evolució de l'aplicació també passaria per millorar la interacció entre usuari i aplicació. És cert que mitjançant el "widget" d'etiquetes es pot seleccionar els països que un vol que apareguin anomenats en el diagrama bivariant. Tanmateix, una millora interessant passaria per permetre a l'usuari, igual que el mapa, rebre informació instantània en passar el ratolí per un punt.

Aquestes millores passen per aprofundir en l'aprenentatge del llenguatge de programació R, que per altra banda cal dir que a partir de la realització s'ha descobert que té un gran potencial i que és una eina realment potent pel que fa al tractament estadístic de grans "sets" de dades.

Avenços de més envergadura podrien ser els següents:

Seria realment útil no restringir l'aplicació a un conjunt de dades escollides pel creador, sinó que seria interessant poder vincular-la a la font d'on s'han extret les dades, en aquest cas la pàgina de dades obertes del Banc Mundial. Això permetria a l'usuari escollir tants indicadors com trobés interessant.

A més a més, una altra millora que proporcionaria un salt de qualitat a l'aplicació seria que tingués un sistema de carregar dades. És a dir, la interfície creada estudia indicadors fins al 2012. En el futur probablement interessarà estudiar els indicadors més recents, per tant seria d'una utilitat notòria el fet de poder carregar les noves dades sense necessitat d'haver

de canviar el codi de programació ni d'haver de descarregar arxius de dades constantment. Això té relació amb el paràgraf anterior i segurament passaria per vincular l'aplicació amb la font de les dades.

9.3. Revisió crítica de la feina feta

Mirant enrere, la realització d'aquest projecte ha estat una tasca més laboriosa del que inicialment s'esperava. En especial destacar les hores dedicades a la realització del codi de programació per a l'aplicació.

Malgrat haver cursat dos assignatures en Python, cosa que podia fer pensar que aprendre a programar amb R seria una tasca senzilla, ha resultat que el llenguatge R no s'hi assembla tant, i ha estat necessari cursar tutorials i sol·licitar ajuda.

Com que realment era difícil de predir l'abast del projecte, en el sentit que es desconeixia si seria possible crear una interfície usant eines estadístiques, ha estat difícil calcular els temps i organitzar-se. És per això que considero que ha faltat temps per dedicar a hores d'aprenentatge de programació amb R. Finalment, la recerca feta va resultar ser d'ajuda i va ajudar a prendre idees que es poguessin intentar desenvolupar de cara a la interfície.

Pel que fa a la memòria, la seva realització ha estat costosa, però s'ha anat fent a mesura que avançava el termini.

Agraïments

Principalment, mostrar el més sincer agraïment al meu tutor, el Lluís Marco Almagro, per l'ajuda que m'ha prestat en la realització d'aquest projecte. M'agradaria destacar la disponibilitat i interès que sempre ha mostrat a l'hora de guiar-me i donar-me un cop de mà. Ha estat un plaer treballar amb ell, el tracte que he rebut per part seva ha estat el millor que podria esperar, i considero que el seu suport ha estat vital per a la realització del projecte.

A més a més, voldria donar les gràcies als meus pares, Joan i Carme. Sempre han estat allà per donar-me una empenta necessària en els moments més difícils d'aquests anys d'estudi universitari, i estic segur que seguirà sent així pels que falten.

Bibliografia

Referències bibliogràfiques

Berners-Lee, T. (2012). *Open Data Institute* . Recuperado el Noviembre de 2014, de <http://opendatainstitute.org/>

Berners-Lee, T. (1994). *W3C*. Recuperado el Noviembre de 2014, de <http://www.w3.org/>

Fundación CTIC. (2003). *Fundación CTIC*. Recuperado el Noviembre de 2014, de <https://www.fundacionctic.org/>

Generalitat de Catalunya. (2011). *Dades Obertes gencat*. Recuperado el noviembre de 2014, de <http://dadesobertes.gencat.cat/ca/>

<http://www.r-project.org/>. (2010). Recuperat el Gener de 2014, de R Project: <http://www.r-project.org/>

Iprocuratio, C. (2010). *Iprocuratio Consultores*. Recuperado el Noviembre de 2014, de <http://www.iprocuratio.com/>

Ministerio de Industria, Energía y Turismo y Ministerio de Hacienda y Administraciones Públicas. (2011). *datos.gob.es/*. Recuperado el noviembre de 2014, de <http://datos.gob.es/>

Observatorio Nacional de las Telecomunicaciones y la Sociedad de la Información. (2012). *Estudio de caracterización del sector infomediario*.

Open Knowledge Foundation. (2004). *Ckan.org*. Recuperado el noviembre de 2014, de <http://ckan.org/>

The World Bank. (1944). *data.worldbank.org*. Recuperado el diciembre de 2014, de <http://data.worldbank.org/>

Universitat Pompeu Fabra. (2012). *data.upf.edu/*. Recuperado el diciembre de 2014, de http://data.upf.edu/en/linked_data

World Wide Web Foundation. (2012). *The Web Index*. Recuperado el diciembre de 2014, de <http://thewebindex.org/data/index/>



ANNEX

En aquest Annex s'incorpora el codi de programació utilitzat per a la realització de la interfície.

Com s'ha explicat prèviament, una aplicació de Shiny conté dos fitxers, anomenats "ui.R" i "server.R". A més a més, aquesta aplicació en concret, igual que moltes d'altres, afegeix un arxiu anomenat "helpers.R", que emmagatzema funcions necessàries per als altres dos arxius i d'aquesta manera evita que s'hagi d'escriure tot el codi inclòs a la funció cadada vegada, a més que facilitat la lectura i comprensió del codi dels arxius "ui.R" i del "server.R".

CODI ui.R

```
library(shiny)
load(file = 'data/world.data.RData')

shinyUI(navbarPage(
  title = 'Explorador socioeconòmic',

  tabPanel('Explorador de dades',
    fluidRow(
      column(4,
        selectizeInput("exp.country",
          "País",
          c("Tots",
            unique(as.character(world.data$country.name))))
      ),
      column(4,
        selectizeInput("exp.year",
          "Any:",
          c("Tots",
            unique(as.character(world.data$year))))
      )
    ),
    fluidRow(
      dataTableOutput('ex3')
    )
  ),

  tabPanel('Anàlisi de relacions',
    sidebarLayout(
      sidebarPanel(
        selectInput("sc.x",
          "Variable X:",
          c("mortalitat infantil", "esperança de vida", "atur", "PIB per
capita")),
        selectInput("sc.y",
          "Variable Y:",
          c("mortalitat infantil", "esperança de vida", "atur", "PIB per
capita"), selected="esperança de vida"),
        selectizeInput("sc.countries",
          "Mostrar etiquetes dels següents països:",
          unique(as.character(world.data$country.name))),
        multiple=TRUE,

        sliderInput("sc.year", "Any:",
          min=1991, max=2012, value=2012, animate=TRUE)
      ),
      mainPanel(
```



```

        plotOutput("sc", width = "100%", height = "800px")
      )
    ),
    tabPanel('Anàlisi geogràfic',
      fluidRow(
        column(2,
          selectInput("map.indicator",
            "Indicador:",
            c("mortalitat infantil", "esperança de vida", "atur", "PIB
per capita"))
        ),
        column(2, offset=1,
          sliderInput("map.year", "Any:",
            min=1991, max=2012, value=2012, animate=TRUE)
        ),
        column(2, offset=1,
          checkboxInput("map.absolut", label = "Colorejar en termes absoluts?",
value = TRUE)
        )
      ),
      fluidRow(
        plotOutput("map", width = "100%", height = "800px")
      )
    )
  )
))

```

CODI server.R

```

library(shiny)
library(rworldmap)
library(RColorBrewer)
library(ggplot2)
library(classInt)

load(file = 'data/world.data.RData')
load(file = 'data/country.codes.RData')

source("helpers.R")

shinyServer(function(input, output) {

  output$sc <- renderPlot({
    args <- list(data=subset(world.data, year==input$sc.year))
    args$x <- switch(input$sc.x,
      "mortalitat infantil" = 5,
      "esperança de vida" = 7,
      "atur" = 6,
      "PIB per capita" = 9
    )
    args$y <- switch(input$sc.y,
      "mortalitat infantil" = 5,
      "esperança de vida" = 7,
      "atur" = 6,
      "PIB per capita" = 9
    )

    args$xlim <- switch(input$sc.x,
      "mortalitat infantil" = c(2, 350),
      "esperança de vida" = c(25, 85),
      "atur" = c(0, 40),

```

```

        "PIB per capita" = c(50, 200000)
    )
    args$ylim <- switch(input$sc.y,
        "mortalitat infantil" = c(2, 350),
        "esperança de vida" = c(25, 85),
        "atur" = c(0, 40),
        "PIB per capita" = c(50, 200000)
    )

    args$xlabel <- input$sc.x
    args$ylabel <- input$sc.y

    args$countries <- input$sc.countries

    do.call(country.sc, args)
})

output$map <- renderPlot({
    subdata <- world.data[world.data$year == input$map.year,]
    args <- switch(input$map.indicator,
        "mortalitat infantil" = list(subdata, "mortalitat.infantil",
brewer.pal(9, "Reds")),
        "esperança de vida" = list(subdata, "esp.vida", brewer.pal(9,
"Greens")),
        "atur" = list(subdata, "atur", brewer.pal(9, "Purples")),
        "PIB per capita" = list(subdata, "PIBpercapita", brewer.pal(9, "Blues"))
    )

    args$absolut <- input$map.absolut

    args$limits <- switch(input$map.indicator,
        "mortalitat infantil" =
classIntervals(world.data$mortalitat.infantil, n=9)[["brks"]],
        "esperança de vida" = classIntervals(world.data$esp.vida,
n=9)[["brks"]],
        "atur" = classIntervals(world.data$atur, n=9)[["brks"]],
        "PIB per capita" = classIntervals(world.data$PIBpercapita,
n=9)[["brks"]])

    do.call(country.map, args)
})

output$ex3 <- renderDataTable({
    data <- world.data
    if (input$exp.country != "Tots"){
        country.tmp <- country.codes[country.codes$country.name==input$exp.country,2]
        data <- data[data$country.code == country.tmp,]
    }
    if (input$exp.year != "Tots"){
        data <- data[data$year == input$exp.year,]
    }
    data
})

})

```



CODI helpers.R

```

country.map <- function(data, ind.column, color, absolut=FALSE, limits) {

  sMapa <- joinCountryData2Map(data, joinCode = "ISO3", nameJoinColumn = "country.code")
  if (!absolut) {
    mapParams <- mapCountryData(sMapa, nameColumnToPlot = ind.column, numCats=9, mapTitle=
  "", addLegend = FALSE, colourPalette = color, catMethod="quantiles")
  }
  if (absolut) {
    mapParams <- mapCountryData(sMapa, nameColumnToPlot = ind.column, numCats=9, mapTitle=
  "", addLegend = FALSE, colourPalette = color, catMethod=limits)
  }

  do.call (addMapLegend, c(mapParams, legendWidth=0.5, legendMar=2))
}

country.sc <- function(data, x, y, xlab, ylab, xlim, ylim, countries="") {

  country.labels <- rep("", times=length(data$country.name))
  country.rows <- pmatch(countries, data$country.name)

  country.labels[country.rows] <- as.character(data$country.name[country.rows])

  theme_set(theme_bw(base_size = 25))
  qplot(data[,x], data[,y], size=data$poblacio, xlim=xlim, ylim=ylim, color=data$region,
  alpha=0.5, xlab=xlab, ylab=ylab)+ scale_size_continuous(range = c(4, 20)) +
  theme(legend.position="none")+ scale_colour_brewer(palette="Set1") +
  geom_text(aes(data[,x], data[,y],label=country.labels), size=6, color="black",
  fontface="bold")
}

```